

## THESIS / THÈSE

### DOCTOR OF SCIENCES

#### Invariant stabilization of discretized boundary control systems

Dehayé, Jonathan

*Award date:*  
2017

*Awarding institution:*  
University of Namur

[Link to publication](#)

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



**UNIVERSITÉ DE NAMUR**

FACULTÉ DES SCIENCES

DÉPARTEMENT DE MATHÉMATIQUE

# **Invariant stabilization of discretized boundary control systems**

Thèse présentée par  
**Jonathan Dehaye**  
pour l'obtention du grade  
de Docteur en Sciences

Composition du Jury:

Anne LEMAÎTRE (Président)  
Alexandre MAUROY  
Christophe PRIEUR  
Alain VANDE WOUWER  
Joseph WINKIN (Promoteur)

Octobre 2017

Graphisme de couverture : © Presses universitaires de Namur

© Presses universitaires de Namur & Jonathan Dehay

Rempart de la Vierge, 13

B-5000 Namur (Belgique)

Toute reproduction d'un extrait quelconque de ce livre,  
hors des limites restrictives prévues par la loi,  
par quelque procédé que ce soit, et notamment par photocopie ou scanner,  
est strictement interdite pour tous pays.

Imprimé en Belgique

ISBN : 978-2-87037-775-8

Dépôt légal: D/2017/1881/42

Université de Namur

Faculté des Sciences

Rue de Bruxelles, 61, B-5000 Namur (Belgique)

## **Stabilisation invariante de systèmes avec commande frontière discrétisés**

par Jonathan Dehay

**Résumé :** Stabilisation et invariance sont les deux mots-clefs de ce travail. Par stabilisation invariante, l'on entend le fait de stabiliser asymptotiquement un système tout en maintenant les trajectoires d'état dans un domaine prédéfini. Dans un premier temps, nous traitons le cas des systèmes positifs linéaires temps-invariants (LTI) fini-dimensionnels, pour lesquels nous discutons de la pertinence du choix d'une entrée positive pour le processus de stabilisation, fournissons une paramétrisation des feedbacks positivement stabilisants pour une certaine classe de systèmes positifs, et étendons le concept d'invariance à des cônes, secteurs ou ensembles de niveau d'une fonction de Lyapunov. Ensuite nous adaptons les résultats aux systèmes positifs LTI infini-dimensionnels, nous expliquons le lien entre une entrée agissant dans les conditions frontières et cette même entrée agissant dans la dynamique, nous appliquons les résultats au modèle de diffusion pure, et discutons des conditions frontières lors de la discrétisation d'un système EDP. Finalement, nous traitons le cas des systèmes positifs non linéaires temps-invariants (NTI) infini-dimensionnels, pour lesquels nous adaptons une nouvelle fois les résultats théoriques précédents et présentons un exemple pertinent, à savoir un modèle de réacteur biochimique.

## **Invariant stabilization of discretized boundary control systems**

by Jonathan Dehay

**Abstract:** Stabilization and invariance are the two keywords of this work. By invariant stabilization, one should understand the asymptotic stabilization of a system while keeping the state trajectories in a predetermined domain. First, we deal with the positive linear time-invariant (LTI) finite-dimensional systems for which we discuss the relevance of choosing a nonnegative input for the stabilization process, we provide a parameterization of all positively stabilizing feedbacks for a particular class of positive systems, and we extend the concept of invariance to cones, sectors and Lyapunov level sets. Then, we adapt the results to the positive LTI infinite-dimensional systems, we explain how one can switch from an input acting in the boundary conditions to an input acting in the dynamics, we introduce the standard example of the pure diffusion, and we discuss the boundary conditions when discretizing a PDE system. Finally, we deal with the positive nonlinear time-invariant (NTI) infinite-dimensional systems, for which we once again adapt the previous theoretical results and consider a relevant example, namely a biochemical reactor model.

Thèse de doctorat en Sciences Mathématiques (Ph.D. thesis in Mathematics)

Date: 25/10/2017

Département de Mathématique

Promoteur (Advisor): Prof. J. WINKIN



## Remerciements

J'ai gardé ce passage pour la fin, mon tout dernier ajout à mon manuscrit de thèse. Ironiquement, il s'agit sans doute de la partie la plus difficile à écrire. Pas parce que je ne sais pas qui remercier - j'ai de nombreux noms en tête et ils sont très clairs - mais parce qu'il n'est jamais facile de décrire avec de simples mots toute l'affection et la reconnaissance que l'on porte à une personne. C'est une étape de plus de six années qui se termine pour moi, et j'ai naturellement beaucoup de personnes à remercier, des personnes qui m'ont accompagné, de près ou de loin, d'une manière ou d'une autre, durant cette aventure.

Je tiens avant tout à remercier mon promoteur, Joseph Winkin, de m'avoir épaulé et encadré pendant ces années de recherche. Je le remercie pour son optimisme, sa bonne humeur, sa combativité et sa méticulosité. Je le remercie pour sa patience, sa disponibilité, son aide et ses conseils innombrables. Je le remercie de m'avoir appris la positivité, tant dans les mathématiques que dans la vie de tous les jours, et je le remercie d'avoir permis l'aboutissement de ce travail. Pour tout ça, Joseph, merci.

Mes remerciements vont également aux membres de mon comité d'accompagnement, Christophe Prieur (*GIPSA-lab, Dpt of Automatic Control, Grenoble*) et Alain Vande Wouwer (*UMons, Control Dpt, Mons*), ainsi qu'à Anne Lemaître et Alexandre Mauroy, qui ont accepté de faire partie de mon jury de thèse. Merci à vous pour vos nombreux conseils et vos remarques constructives, lors de l'épreuve de confirmation et lors de la défense privée, qui ont contribué à l'amélioration du manuscrit, tant sur la forme que sur le fond.

J'ai la chance d'avoir une famille exceptionnelle, sans laquelle je ne serais pas là pour écrire ces lignes. A Jérémy, mon binôme depuis presque trente ans, qui est avant tout le frère, mais qui a aussi été le colocataire, l'assistant, le collègue de bureau, et bien sûr le pote, le pote qui est là quand on rigole, mais qui est là aussi quand on a besoin d'un coup de main... le vrai Pote, avec un "P" majuscule s'il-vous-plait. A Pauline, qui me soutient (et me supporte) depuis presque huit ans, avec qui j'ai passé de nombreuses grandes étapes de la vie, avec qui j'ai affronté de nombreux défis, le dernier en date étant l'achat de la maison (et l'adoption d'une petite boule de poils "légèrement turbulente" nommée Murphy), et avec qui je compte encore en affronter bien d'autres. A mes parents, les meilleurs qu'on puisse trouver dans l'univers (si si, je peux le prouver mathématiquement), qui nous ont toujours soutenus, mon frère et moi, dans nos décisions (même celle de faire des études de mathématiques) et qui se plient en quatre depuis notre plus jeune âge pour que nous puissions profiter au mieux de la vie. A Papy Michel et Mamy Marie-Thérèse (belote, films et tranches de rire à n'en plus finir pour une complicité encore inégalée), Papy Hector, Fabrice, Taty, et les cousins (petits sagouins!) bien sûr, Adrien et Lisa. A ma belle-famille, Francis, Michèle, Céline, Jean-Pol, Colette, Val et Sam, pour tous les coups de main, et pour le contact qui passe si bien. A vous tous, merci, merci, et mille fois merci.

Il est toujours plus facile - et plus marrant - d'avancer dans la vie quand on est bien entouré. On dit que, contrairement à sa famille, on choisit ses amis... Au vu des miens, on pourrait se poser de sérieuses questions quant à mes critères de sélection... J'aimerais ainsi remercier la crème de la crème : Jérémy (une deuxième fois, trop de

casquettes), Martin "Martinez" Gueuning, William "Willy" Henrotin, Jérémy "Watson" Rekier, Xavier "Maximus" Toledo, Laurent "Plouf" Standaert, Nicolas "Gronaze" Mourmaux, Adrien "Boubou" Rogé et Margo "la Mouche" Genette, Charles "Lipschitz" Hubaux et Eve Tilman, Johan "Jojo" Barthelemy, Sébastien "Chabal" Canart, Xavier "Myllas" Lambinon, Maxime "Max" Fondaire, Clément "Jinem" Desquilbet, Bruno "Caillak" Scaillet, Kevin "Hazrae" Voneche, Vivian "Crouton Nerveux" Antoine, Julien "Raisin" Christophe, François "Fradz" Lutz, Samuel "Sammy" Warin, Laurie "Lolo" Hollaert et Antoine "Toutoune" Pecceu, Lionel "Ruth" Ruth et Marie "Fifille" Haufroid, Laurent "Tack Os" Demez, Frédéric "Ambi" Ambroise, Jérôme "Jey" Gillet, François "Gauth" Braun, Denis "le Troll" Noël, Céline "Marraine" Renkens, Marie Moriamé, Dimitri "Dim" Moriamé, Sandy Vautron, Florian "Flo" Baetmans, Claire "FSP" Linart, Charlotte Tannier, Maud Henrotte, Julie Scheen... A vous tous, un énorme merci. La vie, c'est plus rigolo quand vous êtes là.

A mes collègues assistants ou chercheurs, anciens comme nouveaux, qui ont été et sont bien plus que des collègues en vérité : Anne-Sophie, Delphine, Manon, Pauline, Eve, Julien et Julien, François et François, Morgane, Marie, Arnaud, Nicolas, et tous les autres qui ont déjà été cités ci-dessus (ne soyez pas trop gourmands)... Merci! Travailler à vos côtés a été un réel plaisir, et rigoler avec vous l'a été tout autant! A mes collègues académiques ou scientifiques avec qui j'ai eu l'occasion de collaborer (Anne, Joseph, Anne-Sophie, Martine, Benoît, Annick, André H., Eric) ou pas (Téo, André F., Valérie, Alexandre, Renaud, Carole, Marcel), à nos deux secrétaires exceptionnelles, Pascale et Alice, à vous tous, un énorme merci pour ces années et cette ambiance formidable.

Last but not least, je souhaite remercier chaleureusement les dix générations d'étudiants que j'ai vu défiler et évoluer pendant mes années d'assistanat. Je n'ai pas la place ici pour citer tous vos noms, mais soyez bien sûrs que je les ai en tête au moment où j'écris ces lignes, et pour de nombreuses années encore (si ma mémoire défaillante me le permet). Vous avez été extras, chacun d'entre vous, et ç'a été un réel plaisir de pouvoir vous donner cours. Je vous souhaite bonne route, quelle qu'elle soit!

A vous tous, merci du fond du coeur pour cette formidable expérience.

*"Aurora, what is love known by?"  
"When it hurts to say goodbye."*

– Aurora & Igniculus

# Contents

<b>Nomenclature</b>	<b>1</b>
<b>Introduction</b>	<b>5</b>
<b>1 Positive linear time-invariant finite-dimensional systems</b>	<b>11</b>
1.1 Framework . . . . .	11
1.2 Positive stabilization by nonnegative input . . . . .	14
1.2.1 A classic approach . . . . .	14
1.2.2 Working on an extended system . . . . .	15
1.3 Designing positively stabilizing feedbacks . . . . .	19
1.4 Invariant stabilization . . . . .	22
1.4.1 Cone invariance . . . . .	22
1.4.2 Sector invariance . . . . .	27
1.4.3 Lyapunov level sets . . . . .	32
<b>2 Positive linear time-invariant PDE systems</b>	<b>37</b>
2.1 Positive stabilization of a positive system by nonnegative input . . . .	37
2.2 Application: the pure diffusion system . . . . .	39
2.2.1 Modeling . . . . .	39
2.2.2 Feedback design . . . . .	45
2.2.3 Numerical simulations . . . . .	54
2.3 Convergence analysis . . . . .	66
2.3.1 Choosing a feedback . . . . .	67
2.3.2 Consistency and stability of the scheme . . . . .	69
2.3.3 State space approach . . . . .	71
2.4 Positive stabilization preserving boundary conditions . . . . .	76
2.4.1 A classic approach . . . . .	77
2.4.2 Application: the pure diffusion system . . . . .	81
<b>3 Positive nonlinear time-invariant PDE systems</b>	<b>89</b>
3.1 From a positive stabilization problem to an invariant stabilization problem . . . . .	89
3.2 Application: a tubular biochemical reactor model . . . . .	90
3.2.1 Description of the system . . . . .	90



3.2.2	Considering a cone . . . . .	101
3.2.3	Considering a sector . . . . .	108
3.2.4	Considering Lyapunov level sets . . . . .	111
<b>Conclusion</b>		<b>119</b>
<b>Bibliography</b>		<b>123</b>

# Nomenclature

## List of notations

$t$	The time variable
$z$	The space variable
$\dot{x}(t)$	The derivative of $x$ with respect to $t$
$x'(z)$	The derivative of $x$ with respect to $z$
$\mathbb{R}_+$	The set of nonnegative real numbers
$\mathbb{R}_{0,+}$	The set of positive real numbers
$\mathbb{C}_+$	The set of complex numbers with nonnegative real parts
$\mathbb{R}_+^n$	The set of nonnegative real vectors of dimension $n$ i.e. vectors with nonnegative real components
$\mathbb{R}_{0,+}^n$	The set of strictly positive real vectors of dimension $n$ i.e. vectors with positive real components
$\text{Re}(c)$	The real part of the complex number $c$
$A \geq 0$	The matrix $A$ is nonnegative (i.e. $a_{ij} \geq 0$ for all $i, j$ )
$v \gg 0$	The vector $v$ is strictly positive (i.e. $v_i > 0$ for all $i$ )
$v > 0$	The vector $v$ is positive (i.e. $v_i \geq 0$ for all $i$ , and $v \neq 0$ )
$v \geq 0$	The vector $v$ is nonnegative (i.e. $v_i \geq 0$ for all $i$ )
$v^T$	The transpose of the vector $v$
$A^T$	The transpose of the matrix $A$
$A^{-1}$	The inverse of the matrix $A$
$\sigma(A)$	The spectrum of the matrix $A$
$\rho(A)$	The resolvent set of the operator $A$
$R(\lambda, A)$	The resolvent operator of $A$ , given by $(\lambda I - A)^{-1}$ with $\lambda \in \rho(A)$
$I$	The identity matrix or operator
$\text{rk}(A)$	The rank of the matrix $A$
$\mathcal{C}$	The controllability matrix of the system

$\mathfrak{c}\{v_1, \dots, v_n\}$	The cone finitely generated by the vectors $v_1, \dots, v_n$
$\mathring{\mathfrak{c}}\{v_1, \dots, v_n\}$	The interior of the cone finitely generated by the vectors $v_1, \dots, v_n$
$e_i$	The $i^{\text{th}}$ vector of the canonical basis
$\text{sgn}(r)$	The sign of the real number $r$
$\det(A)$	The determinant of the matrix $A$
$A^{[i]}$	The matrix $A$ with the $i^{\text{th}}$ row and column removed
$\mathring{K}$	The interior of the cone $K$
$K_0$	The cone $K$ without the origin
$r(A)$	The spectral radius of the operator $A$
$s(A)$	The spectral bound of the operator $A$
$\omega_0(T)$	The growth bound of the $C_0$ -semigroup $(T(t))_{t \geq 0}$
$A^*$	The adjoint of the operator $A$
$\langle v, w \rangle$	The inner product between $v$ and $w$
$\delta_{ij}$	The Kronecker delta
$\delta_0$	The Dirac delta distribution at $z = 0$
$\mathcal{D}(A)$	The domain of the operator $A$
$C(i, j)$	The binomial coefficient indexed by $i$ and $j$
$L^p(a, b)$	The Lebesgue space of complex-valued $p$ -integrable functions defined on $[a, b]$
$H^p(a, b)$	The Sobolev space of order $p$
$\mathcal{L}(X, Y)$	The vector space of all bounded linear operators from $X$ to $Y$
$\chi_S$	The characteristic function of the set $S$
$\ker(L)$	The kernel of the linear map $L$
$A_l^\#$	The left pseudoinverse of the matrix $A$

**List of abbreviations**

a.c.	absolutely continuous
LMI	linear matrix inequality
LTI	linear time-invariant
NTI	nonlinear time-invariant
ODE	ordinary differential equation
PDE	partial differential equation
vs.	versus
w.l.o.g.	without loss of generality



# Introduction

*We see the world as a mystery,  
a puzzle, because we've always  
been a species of problem-solvers.*

– Alexandra Drennan

Control is something with which each of us is confronted on a daily basis, often without even realizing it. It is a fundamental aspect of the everyday life, as one can actually observe control processes everywhere, for example in cars (steering, thermostats and air conditioning units, cruise control...), airplanes (flight path / air traffic control systems using GPS...), robotic arms, segways, drones, nuclear or biochemical reactors, and even in our own human bodies, as for example the perspiration process, the expansion or contraction of the pupil depending on the ambient luminosity, the elimination of toxins or medicines in the blood system, and many others. The concept of controlling in order to regulate, stabilize, or, on a more general level, to force physical systems to adopt a specific desired behavior, may be essential with a view to maximizing a production, avoiding undesired behaviors that could lead to wrong functioning or material degradation, minimizing the loss of resources or the waste generation...

In this work we deal with positive linear systems, which are linear systems whose state variables are nonnegative at all time. Studying this kind of systems is of great importance as the nonnegativity property can be found frequently in numerous fields like biology, chemistry, physics, ecology, economy or sociology (see e.g. (Abouzaid, Winkin and Wertz 2010), (Chellaboina, Bhat, Haddad and Bernstein 2009), (Dehay 2011), (Dehay and Winkin 2016), (Dehay and Winkin 2017), (Farina and Rinaldi 2000), (Haddad, Chellaboina and Hui 2010), (Laabissi, Achhab, Winkin and Dochain 2001), (Winkin, Dochain and Ligarius 2000) for particular examples). Although the mathematical models preserve the positivity property of the state, controlling the system might lead to a violation of the property. So when stabilizing a system, one cannot force a state variable representing e.g. a mass, a density or a concentration to become negative at some time in order to make it asymptotically stable, as this would make no sense on a physical point of view and would lead to unpractical and meaningless results. It is then essential to force the nonnegativity of the state at all time when

studying positive systems to ensure that the very nature of the system is preserved and that the mathematical methodology makes sense from the point of view of the applications. In this work, the notion of positivity is generalized to the notion of invariance, for which we do not only consider state variables that can only be nonnegative, but also state variables that are restricted to a cone, i.e. state variables that have lower bounds or upper bounds exclusively, or to a sector, i.e. state variables that have lower and upper bounds at the same time. That kind of restriction could happen naturally, e.g. in systems that hold speeds or temperatures as state variables, or could be forced manually to avoid undesired configurations due to technology limitations, financial restrictions, etc. It is then essential to extend the results to this larger class of systems.

The main matter of this work is precisely to mix both concepts of stabilization and invariance for a particular type of systems: we deal with the invariant stabilization of time-invariant differential systems, more specifically discretized boundary control systems. By invariant stabilization, one should understand the asymptotic stabilization of a system while keeping the state trajectories in a predetermined domain. This is far from being a trivial objective, and one often has to face compromises to reach this goal.

Positive (and invariant) stabilization is a topic that has been studied for many years now (see e.g. (Berman, Neumann and Stern 1989) for nonnegative matrices theory in positive differential and control systems, (Boyd, El Ghaoui, Feron and Balakrishnan 1994) for LMIs in positive orthant stabilizability, (Castelan and Hennes 1993) for positively invariant polyhedra for linear continuous-time systems, (Hartman 1972) for the invariance of closed sets, (Saperstone 1973) for controllability with positive controls) and that is fast-growing since the 2000s (see e.g. (Abouzaid et al. 2010) and (Achhab and Winkin 2014) for the positivity and positive stabilization of linear infinite-dimensional systems, (Bátkai, Fijavž and Rhandi 2017) for positive operator semigroups, (Blanchini 1999) and (Blanchini and Miani 2007) for invariant sets and Lyapunov functions, (Chellaboina et al. 2009) for positivity in systems of chemical reactions, (De Leenheer and Aeyels 2001) for the stabilization of positive linear systems, (Farina and Rinaldi 2000) for positive linear systems analysis, (Haddad et al. 2010) for positive and compartmental systems, (Kaczorek 2002) for positive 1D and 2D, discrete-time and continuous-time systems, (Karafyllis and Krstić 2016) for the positive stabilization of an age-structured chemostat model by use notably of a logarithmic transformation, (Roszak and Davison 2009) for necessary and sufficient conditions for the stabilization of positive LTI systems using a vertex algorithmic approach, (Zhang and Prieur 2017) for positive hyperbolic systems and Lyapunov functions). This is a hot topic nowadays with numerous applications in engineering, pharmacokinetics, biochemistry, ecology and many other fields, and since 2003, the International Symposium on Positive Systems (POSTA) has taken place regularly, putting the spotlight on a wide range of topics from the theory and applications of positive systems. This work brings several contributions in the field of systems and control, in particular to invariant stabilization problems. One of its strong points is that it includes, at the same time, original theoretical results with proofs, relevant and interesting applications, namely the pure diffusion system and a tubular biochemical reactor model, and

numerical simulations with self-developed and optimized MATLAB routines. We have also tried to make this thesis as self-contained as possible, notably by adding a few relevant theoretical developments borrowed from the literature.

The manuscript is organized as follows.

Chapter 1 deals with positive linear time-invariant finite-dimensional systems.

In Section 1.1 we provide some necessary tools, among which definitions, theorems and geometrical interpretations, for a clear understanding of the problem.

In Section 1.2 it is shown in two particular ways why one should not consider a non-negative input in order to positively stabilize a positive system.

In Section 1.3 we provide a systematic way to parameterize all positively stabilizing feedbacks for a particular class of positive systems.

In Section 1.4 we show how one can extend the concept of positivity to a more general concept that is invariance, positivity being a particular case of cone invariance. We consider the invariance of cones, sectors, and ellipsoidal sets that correspond to Lyapunov functions level sets.

Chapter 2 focuses on positive linear time-invariant infinite-dimensional systems.

In Section 2.1 we show that, ironically and unlike what is stated in Section 1.2, it is actually possible to positively stabilize a positive system while considering a non-negative input. However, there is a trick as the input changes whether it acts in the boundary conditions or in the dynamics.

In Section 2.2 we introduce a standard example, namely the pure diffusion system, to support the theoretical results and to perform numerical simulations.

In Section 2.3 we study the convergence of the discretized closed-loop system to the closed-loop PDE system by analyzing the discretization scheme and by means of a state space approach.

In Section 2.4 we raise the question of the boundary conditions status when discretizing a PDE system, and we provide a method to ensure that the discretized boundary conditions are always verified for the finite-dimensional model. The results are supported by the pure diffusion system, once again.

Chapter 3 introduces positive nonlinear time-invariant infinite-dimensional systems.

In Section 3.1 it is shown how the positive stabilization problem of a positive nonlinear system can be seen as an invariant stabilization problem of an invariant (more precisely, a cone invariant) linear system. This allows to apply the previous theoretical results, and leads to a simpler though relevant problem.

In Section 3.2 we study a relevant example, namely a tubular biochemical reactor model, and bring a solution to the invariant stabilization problem for that particular case of positive nonlinear system.

We close this work by giving some concluding remarks and providing some interesting perspectives for future work.



## Contributions

The main contributions include:

- formal proofs, completing (De Leenheer and Aeyels 2001, Theorem 4) which is inspired by the results in (Saperstone 1973), that show that positive stabilization of a positive LTI finite-dimensional system by nonnegative input cannot be achieved, one following a classic approach (considering the state trajectory) and the other following a different, more peculiar approach (considering an extended system), and the extension of the result to positive LTI infinite-dimensional systems;
- a study of how the input sign can change depending on whether it acts in the boundary conditions or in the dynamics (see (Emirsjlow and Townley 2000)), implying that the positive stabilization of a positive system by nonnegative input is still possible, in a sense;
- the design of all positively stabilizing feedbacks for a particular class of positive systems brought as a theoretical result with a formal proof, along with a systematic way (involving the positive stabilization theory, see e.g. (Chellaboina et al. 2009), (Haddad et al. 2010), (Roszak and Davison 2009), and linear programming, see (Schrijver 1998)) to fully parameterize these feedbacks;
- a deep analysis of a standard and classic example, namely the pure diffusion system, for which we design a discretized model of the system, perform numerical simulations, and provide a convergence study considering the discretization scheme (see e.g. (Frey 2009), (Thomas 1995)) and a state space approach (see (Emirsjlow and Townley 2000));
- results on how to preserve boundary conditions for discretized systems, as boundary conditions are hardly ever considered when it comes to discretized systems: it is shown how to achieve positive stabilization of a positive system while preserving both positivity and boundary conditions, using the polyhedral sets invariance theory (see e.g. (Castelan and Hennet 1993)), and the results are applied to the pure diffusion model;
- an extension of positivity to the invariance of a cone or a sector, and in particular an extension of the results concerning the invariance of a cone (see (Beauthier 2011)) to the invariance of a sector, for which multiple theoretical results are given with proofs;
- the invariance study by Lyapunov level sets: considering a closed-loop system and using Lyapunov functions, it is shown in an original and efficient way how to design a set of initial conditions that ensure that the state remains in a given cone or sector (see (Blanchini and Miani 2007, Chapter 4)), including a theoretical result with proof and a strong geometrical background;

- a deep analysis of a relevant and interesting example, namely a biochemical reactor distributed parameter model (see (Dramé, Dochain and Winkin 2008)), for which the previous theoretical results are applied to a linearized and discretized model of the system;
- many routines that have been coded in MATLAB and optimized to support the theoretical results and to provide a thorough analysis of both studied applications.

A detailed list enumerating our most important communications and publications can be found below, whose contributions are included in this thesis.

- J.N. Dehaye and J.J. Winkin, "Parameterization of positively stabilizing feedbacks for single-input positive systems" in *Systems & Control Letters* 98, 57-64 (2016)
- J.N. Dehaye and J.J. Winkin, "Positive stabilization of a diffusion system by nonnegative boundary control" in *Positive Systems, Theory and Applications (POSTA 2016)*, Lecture Notes in Control and Information Sciences Vol. 471, 179-190 (2017)
- "Parameterization of Positively Stabilizing Feedbacks for Single-Input Positive Systems" at the SIAM Conference on Control and Its Applications, 8-10 July 2015 in Paris, France
- "Positive stabilization of a diffusion system by nonnegative boundary control" at the 5<sup>th</sup> International Symposium on Positive Systems (POSTA), 14-16 September 2016 in Rome, Italy
- Presentation of a group project "Some positivity preserving schemes for semi-linear problems" at the 15<sup>th</sup> Internet Seminar on Operator Semigroups for Numerical Analysis (ISEM) workshop, 3-9 June 2012 in Blaubeuren, Germany



# Chapter 1

## Positive linear time-invariant finite-dimensional systems

### 1.1 Framework

Before getting to the heart of the matter, it is desirable that we provide the reader with some important results and definitions to ensure that they have the tools needed for a clear understanding of the problem and the solutions brought in this work.

Let us consider a linear time-invariant (LTI) finite-dimensional system described by the equations

$$\begin{cases} \dot{x} &= Ax + Bu \\ y &= Cx + Du \end{cases} \quad (1.1.1)$$

where  $A \in \mathbb{R}^{n \times n}$ ,  $B \in \mathbb{R}^{n \times m}$ ,  $C \in \mathbb{R}^{p \times n}$ ,  $D \in \mathbb{R}^{p \times m}$ ,  $x$  is the state,  $u$  is the input and  $y$  is the output. First we provide the reader with the concept of positive linear system (see e.g. (Farina and Rinaldi 2000), (Haddad et al. 2010), (Kaczorek 2002), (Roszak and Davison 2009)).

**Definition 1** *A linear system (1.1.1) is positive if for every nonnegative initial state  $x_0 \in \mathbb{R}_+^n$  and for every admissible nonnegative input  $u$  (i.e. every piecewise continuous function  $u : \mathbb{R}_+ \rightarrow \mathbb{R}_+^m$ ) the state trajectory  $x$  of the system and the output trajectory  $y$  are nonnegative, i.e. for all  $t \geq 0$ ,  $x(t) \in \mathbb{R}_+^n$  and  $y(t) \in \mathbb{R}_+^p$ .*

Considering a more geometric perspective, one could express the positivity of a system in the following way: if the initial state is in the nonnegative orthant, the state trajectory remains in the nonnegative orthant at all time (see Figure 1.1).

It is possible to express the positivity of a system by use of the matrices  $A$ ,  $B$ ,  $C$  and  $D$  only. In order to do so, two concepts are needed.

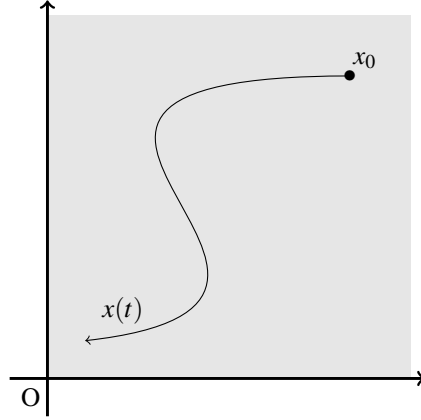


Figure 1.1: The state trajectory remaining in the nonnegative orthant

**Definition 2** A nonnegative matrix  $A$  (denoted by  $A \geq 0$ ) has all its entries greater than or equal to zero, i.e.  $a_{ij} \in \mathbb{R}_+$ , for all  $i, j$ .

**Definition 3** A Metzler matrix  $A$  is a matrix with all its off-diagonal entries greater than or equal to zero, i.e.  $a_{ij} \in \mathbb{R}_+$ , for all  $i \neq j$ .

One can see that a nonnegative matrix is a Metzler matrix, and that a Metzler matrix is a nonnegative matrix up to a diagonal shift. We can now state the previously mentioned result (see e.g. (Farina and Rinaldi 2000), (Haddad et al. 2010)). For a complete proof, we suggest that the reader checks (Dehay 2011, Theorem 1.1).

**Theorem 1.1.1** A linear system (1.1.1) is positive if and only if  $A$  is a Metzler matrix and  $B, C$  and  $D$  are nonnegative matrices.

Before we introduce the notion of positive stabilizability of positive systems, one last concept is needed.

**Definition 4** A matrix  $A$  is (exponentially) stable if there exist positive constants  $M$  and  $\omega$  such that for all  $t \geq 0$

$$\|e^{At}\| \leq Me^{-\omega t}$$

or, equivalently,  $A$  has all its eigenvalues with negative real parts, i.e.  $\text{Re}(\lambda) < 0$ , for all  $\lambda \in \sigma(A)$ .

For convenience, throughout this work the notion of stability will refer to asymptotic stability, which is equivalent to exponential stability as long as we deal with LTI systems.

**Definition 5** A positive LTI system (1.1.1) is positively (exponentially) stabilizable if there exists a state feedback matrix  $K \in \mathbb{R}^{m \times n}$  such that  $A + BK$  is a stable Metzler matrix, i.e. such that there exist positive constants  $M$  and  $\omega$  such that for all  $t \geq 0$

$$\left\| e^{(A+BK)t} \right\| \leq M e^{-\omega t},$$

and for all  $t \geq 0$ ,  $e^{(A+BK)t} \geq 0$ . Such a feedback matrix  $K$  is called a positively stabilizing feedback for the system (1.1.1).

**Remark 1** In Definition 5, it is mentioned that for all  $t \geq 0$ ,  $e^{(A+BK)t} \geq 0$ . This is indeed required for the nonnegativity of the closed-loop trajectory as  $x(t) = e^{(A+BK)t} x_0$  with an arbitrary nonnegative initial state  $x_0$ .

The positive stabilization problem is concerned with existence conditions and the computation of such a matrix  $K$ . This is of great importance as the stabilization process alone does not take the state domain into account. As stated before, positivity is a natural property of many systems and applications. To ensure the consistency of the model and the stabilization process, one should make sure while designing the feedback matrix that the state positivity property is verified at all time.

Now let us introduce the following lemma (see e.g. (Chellaboina et al. 2009, Theorem 5)) that brings the concept of Lyapunov function which will be required in many forthcoming results.

**Lemma 1.1.2** Let a system be described by the equation  $\dot{x} = Ax$  with  $x(0) = x_0$ . Let  $\mathcal{U}$  be an invariant subset with respect to this system, i.e.  $x_0 \in \mathcal{U}$  implies that  $x(t) \in \mathcal{U}$  for all  $t \geq 0$ . Let  $x_e$  be an equilibrium of the system. Let  $V : \mathcal{U} \rightarrow \mathbb{R}$  be a continuously differentiable function and assume that  $V(x_e) = 0$ ,  $V(x) > 0$  for all  $x \in \mathcal{U} \setminus \{x_e\}$ , and  $\dot{V}(x) \leq 0$  for all  $x \in \mathcal{U}$ . Then  $x_e$  is Lyapunov stable with respect to  $\mathcal{U}$ , i.e. for every  $\varepsilon > 0$  there exists  $\delta > 0$  such that, if  $\|x_0 - x_e\| < \delta$ , then  $\|x(t) - x_e\| < \varepsilon$  for every  $t \geq 0$ . If, in addition,  $\dot{V}(x) < 0$  for all  $x \in \mathcal{U} \setminus \{x_e\}$ , then  $x_e$  is asymptotically stable with respect to  $\mathcal{U}$ .

Finally, we state and prove an important and very useful result (see e.g. (Chellaboina et al. 2009), (Haddad et al. 2010), (Roszak and Davison 2009)) which provides a necessary and sufficient algebraic condition for the stability of a Metzler matrix. The proof is inspired by (Haddad et al. 2010, Lemma 2.2), (Rüffer 2010, Lemma 1.1).

**Proposition 1.1.3** A Metzler matrix  $A \in \mathbb{R}^{n \times n}$  is stable if and only if there exists a vector  $v$  in  $\mathbb{R}_{0,+}^n$  such that  $Av$  is in  $\mathbb{R}_{0,-}^n$ .

**Remark 2** Such a real vector  $v$  is said to be a strictly positive vector, and the standard notation  $v \gg 0$  that can often be found in the literature will be used throughout this work. The same goes for strictly negative vectors that we will denote by  $v \ll 0$ . Moreover, a positive vector  $v$  (denoted by  $v > 0$ ) is a non-null vector with nonnegative entries, and a nonnegative vector  $v$  (denoted by  $v \geq 0$ ) is a vector with nonnegative entries. These notations and terminologies also apply to matrices.

**Proof.** Suppose that there exists a vector  $v \gg 0$  such that  $Av \ll 0$ . Now, regarding Lemma 1.1.2, consider the Lyapunov function  $V(x) = v^T p$  where  $p \gg 0$  is solution of the adjoint system described by  $\dot{p} = A^T p$ . Its derivative along the trajectories of the dual system is given by  $\dot{V}(x) = v^T A^T p = (Av)^T p < 0$  for all  $p \gg 0$ . Then  $A^T$  is stable, and so is  $A$ .

Now suppose that  $A$  is stable. As it is Metzler, the resolvent  $R(\lambda, A)$  is positive for every  $\lambda \geq 0$ . In particular,  $R(0, A) = -A^{-1}$  is positive. Then for all  $\tau \gg 0$ ,  $v := -A^{-1}\tau$  is a positive vector such that  $Av \ll 0$ .  $\square$

## 1.2 Positive stabilization by nonnegative input

As stated in the previous chapter, the positive stabilization problem is concerned with existence conditions and the computation of a feedback matrix  $K$  that ensures both the stability of the closed-loop system and the nonnegativity of the state at all time. One obvious way to ensure the nonnegativity of the state trajectory of a positive system is to choose any nonnegative initial condition and to force the input  $u$  to remain nonnegative. However, one can show that it is impossible to positively stabilize an unstable positive system with such an input.

### 1.2.1 A classic approach

Before stating the main result, let us recall the Perron-Frobenius theorem for Metzler matrices (see e.g. (Arrow 1989, Theorem 4), (Horn and Johnson 1990, Theorem 8.4.4)):

**Theorem 1.2.1** *If  $A$  is a Metzler matrix, there exist a real number  $\lambda$  and a real vector  $v > 0$  such that  $Av = \lambda v$ , hence  $\lambda$  is an eigenvalue of  $A$ , and for every eigenvalue  $\mu$  of  $A$ ,  $\operatorname{Re}(\mu) \leq \lambda$ .*

In other words, a Metzler matrix has a real dominant eigenvalue. Note that the result in (Horn and Johnson 1990) is actually shown for nonnegative matrices. However, as stated in the previous section, a Metzler matrix is a nonnegative matrix with a diagonal shift. It is easy to see that a diagonal shift just shifts the eigenvalues and lets the eigenvectors unaffected. Indeed, consider a matrix  $A$  and its shifted matrix  $A_\gamma = A + \gamma I$  with  $\gamma \in \mathbb{R}$ . We then have  $A_\gamma v = (\lambda + \gamma)v$ , which shows that the eigenvalues  $\lambda$  of  $A$  have been shifted from  $\gamma$  and the eigenvectors  $v$  are the same, and thus the result is also valid for Metzler matrices.

In (De Leenheer and Aeyels 2001) it is stated without proof that, in view of (Saperstone 1973), if the dominant eigenvalue of  $A$  is in  $\mathbb{R}_+$  one cannot stabilize the system with a nonnegative input. Then one can conclude that if a positive system is not already stable, it cannot be stabilized by use of a nonnegative input. For the sake of self-containedness, let us briefly formulate and prove that assertion.

**Theorem 1.2.2** *Let the positive linear system described by the equation  $\dot{x} = Ax + bu$ . The system is (exponentially) positively stabilizable by a state feedback  $u = Kx$  such that  $u \in \mathbb{R}_+$  if and only if it is already (exponentially) stable.*

**Proof.** The sufficiency of the condition is trivial as it suffices to take  $K = 0$ , hence  $u = 0$ . Now let us prove the necessity. Suppose that the system is unstable, then by Theorem 1.2.1 the dominant eigenvalue  $\lambda$  of  $A^T$  is in  $\mathbb{R}_+$ , and there exists a positive eigenvector  $v > 0$  such that  $A^T v = \lambda v$ . Now let us define  $\rho = v^T x$  and focus on the unstable part of the system relative to  $\lambda$ . We then have

$$\begin{aligned}\dot{\rho} &= v^T \dot{x} \\ &= v^T Ax + v^T bu \\ &= (A^T v)^T x + v^T bu \\ &= \lambda \rho + (v^T b)u\end{aligned}$$

where  $v^T b \geq 0$ . If the system was positively stabilizable by a state feedback  $u = Kx$  such that  $u \in \mathbb{R}_+$ , then

$$\rho(t) = e^{\lambda t} \rho_0 + \int_0^t e^{\lambda(t-\tau)} (v^T b) u(\tau) d\tau$$

would not tend to zero as  $t$  tends to infinity, since  $\lambda$ ,  $e^{\lambda t}$ ,  $\rho_0$ ,  $v^T b$  and  $u$  are all nonnegative, thus showing that the system cannot be positively stabilized in this way.  $\square$

This result is actually intuitive: as the system is unstable, the state trajectory will naturally not converge to zero. Adding a nonnegative input to the system would make things even worse. This is something one has to take into account when designing positively stabilizing feedbacks, and that illustrates pretty well the fact that reconciling stabilization and positivity is a matter of compromise.

## 1.2.2 Working on an extended system

As we showed the issue of considering a nonnegative input, we try another approach. What we want to do now is to design an equivalent description of the system that guarantees the nonnegativity of the original input while allowing the new input to be without any sign restriction. Consider the system described by the equations

$$\begin{cases} \dot{x} &= Ax + Bu \\ \dot{u} &= v \end{cases}$$

where  $A$  is a Metzler matrix,  $B$  is a nonnegative matrix,  $u$  is the original input and  $v$  is the new input. This leads to the positive extended system described by the equation

$$\begin{pmatrix} \dot{x} \\ \dot{u} \end{pmatrix} = \begin{pmatrix} A & B \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ u \end{pmatrix} + \begin{pmatrix} 0 \\ I \end{pmatrix} v$$



that we will denote by  $\dot{\tilde{x}} = \tilde{A}\tilde{x} + \tilde{B}v$  with initial condition

$$\tilde{x}_0 = \begin{pmatrix} x(0) \\ u(0) \end{pmatrix} = \begin{pmatrix} x_0 \\ u_0 \end{pmatrix} \geq 0$$

and with state feedback control

$$\begin{aligned} v &= \tilde{K}\tilde{x} \\ &= \begin{pmatrix} K_x & K_u \end{pmatrix} \begin{pmatrix} x \\ u \end{pmatrix} \\ &= K_x x + K_u u \end{aligned}$$

where the new input  $v$  has no sign restriction as it represents the variation of the nominal input  $u$ , which allows us - maybe? - to get rid of the input nonnegativity problem. The resulting closed-loop extended system is therefore described by the equation

$$\begin{aligned} \begin{pmatrix} \dot{x} \\ \dot{u} \end{pmatrix} &= \begin{pmatrix} A & B \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ u \end{pmatrix} + \begin{pmatrix} 0 \\ I \end{pmatrix} \begin{pmatrix} K_x & K_u \end{pmatrix} \begin{pmatrix} x \\ u \end{pmatrix} \\ &= \begin{pmatrix} A & B \\ K_x & K_u \end{pmatrix} \begin{pmatrix} x \\ u \end{pmatrix}. \end{aligned}$$

Note that if one considers a static feedback  $v = \tilde{K}\tilde{x}$  for the extended system, it actually corresponds to a dynamic feedback controller  $\dot{u} = K_u u + K_x x$  for the initial system. The extended system is positively stabilizable if and only if there exists a state feedback matrix  $\tilde{K} = \begin{pmatrix} K_x & K_u \end{pmatrix}$  such that

1. the matrix  $\begin{pmatrix} A & B \\ K_x & K_u \end{pmatrix}$  is Metzler, i.e.  $K_u$  is Metzler and  $K_x \geq 0$ , and
2. the matrix  $\begin{pmatrix} A & B \\ K_x & K_u \end{pmatrix}$  is exponentially stable.

We thus have to find  $\tilde{K}$  such that both conditions are verified. However, it is possible to show that there exists  $\tilde{K}$  such that the second one is verified as long as the initial system is stabilizable. Let us introduce the following result.

**Theorem 1.2.3** *The pair  $(A, B)$  is completely controllable (respectively stabilizable) if and only if the pair  $(\tilde{A}, \tilde{B})$  is completely controllable (respectively stabilizable).*

**Proof.** Suppose that the pair  $(A, B)$  is completely controllable. Then  $\text{rk}(\mathcal{C}) = n$ , where  $\mathcal{C}$  is the controllability matrix which is defined by

$$\mathcal{C} = \begin{pmatrix} B & AB & A^2B & \dots & A^{n-1}B \end{pmatrix}.$$

Computing the controllability matrix for the extended system, we obtain

$$\mathcal{C} = \begin{pmatrix} 0 & B & AB & A^2B & \cdots & A^{n-1}B \\ I & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

which is full rank as  $\mathcal{C}$  is full rank and so is the identity matrix. Now suppose that the pair  $(\tilde{A}, \tilde{B})$  is completely controllable. Then  $\mathcal{C}$  is full rank, and as the identity matrix is also full rank, we can deduce that  $\mathcal{C}$  is full rank.

The equivalence of the stabilizability of both the nominal and the extended systems is shown in a similar way. Suppose that the pair  $(A, B)$  is stabilizable. Then  $\text{rk}(sI - A \quad B) = n$ , for all  $s \in \mathbb{C}_+$ . This implies that  $\text{rk}(sI - A \quad -B) = n$  and thus

$$\begin{aligned} \text{rk} \left( sI - \begin{pmatrix} A & B \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 \\ I \end{pmatrix} \right) &= \text{rk} \left( \begin{pmatrix} sI - A & -B \\ 0 & sI \end{pmatrix} \quad \begin{pmatrix} 0 \\ I \end{pmatrix} \right) \\ &= m + n \end{aligned}$$

as the matrix is block triangular which means its rank is equal to the sum of the diagonal block ranks. Now suppose that the pair  $(\tilde{A}, \tilde{B})$  is stabilizable. Then

$$\text{rk} \left( \begin{pmatrix} sI - A & -B \\ 0 & sI \end{pmatrix} \quad \begin{pmatrix} 0 \\ I \end{pmatrix} \right) = m + n,$$

and as the second diagonal block is the identity matrix ( $m \times m$ ) which is full rank, we can deduce that  $\text{rk}(sI - A \quad -B) = n$  and thus  $\text{rk}(sI - A \quad B) = n$ .  $\square$

So, as a consequence the pair  $(\tilde{A}, \tilde{B})$  should be exponentially stabilizable. Now, (Boyd et al. 1994, Section 10.3) provides necessary and sufficient conditions for the positive stabilizability of a positive system, using linear matrix inequalities (LMIs) and a Lyapunov inequality. Let us adapt this result to the extended system, leading to the following theorem.

**Theorem 1.2.4** *Consider a linear time-invariant system described by the equation  $\dot{x} = Ax + Bu$  and its extended system described by  $\dot{\tilde{x}} = \tilde{A}\tilde{x} + \tilde{B}v$  as defined above. The extended system is positively stabilizable if and only if there exist a positive definite block diagonal matrix*

$$Q = \begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix}$$

*and a feedback matrix  $\tilde{K}$  such that, with  $Y = (Y_1 \quad Y_2) = \tilde{K}Q$ , the matrix*

$$\begin{pmatrix} Q_1 A^T + A Q_1 & Y_1^T + B Q_2 \\ Q_2 B^T + Y_1 & Y_2^T + Y_2 \end{pmatrix}$$

*is negative definite,  $Y_1$  is nonnegative and  $Y_2$  is Metzler.*

**Proof.** The extended system is positively stabilizable if and only if (see e.g. (Boyd et al. 1994, Section 10.3)) there exist a positive definite diagonal matrix  $Q$  and a state feedback matrix  $\tilde{K}$  such that, with  $Y = \tilde{K}Q$ ,  $(\tilde{A}Q + \tilde{B}Y)$  is a Metzler matrix and  $Q\tilde{A}^T + Y^T\tilde{B}^T + \tilde{A}Q + \tilde{B}Y$  is negative definite. One easily sees that  $(\tilde{A}Q + \tilde{B}Y)$  is Metzler if and only if the matrix

$$\begin{pmatrix} A & B \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 \\ I \end{pmatrix} \begin{pmatrix} K_x & K_u \end{pmatrix} = \begin{pmatrix} A & B \\ K_x & K_u \end{pmatrix}$$

is Metzler, which means (as stated previously) that  $K_x$  has to be nonnegative and  $K_u$  has to be Metzler. Moreover, as  $Y = \tilde{K}Q$ ,

$$\begin{pmatrix} Y_1 & Y_2 \end{pmatrix} = \begin{pmatrix} K_x & K_u \end{pmatrix} \begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix}$$

and then

$$\begin{cases} K_x &= Y_1 Q_1^{-1} \\ K_u &= Y_2 Q_2^{-1} \end{cases}$$

which implies that  $Y_1$  has to be nonnegative and  $Y_2$  has to be Metzler. Now, we can rewrite  $Q\tilde{A}^T + Y^T\tilde{B}^T + \tilde{A}Q + \tilde{B}Y$  as

$$\begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix} \begin{pmatrix} A^T & 0 \\ B^T & 0 \end{pmatrix} + \begin{pmatrix} Y_1^T \\ Y_2^T \end{pmatrix} \begin{pmatrix} 0 & I \end{pmatrix} + \begin{pmatrix} A & B \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix} + \begin{pmatrix} 0 \\ I \end{pmatrix} \begin{pmatrix} Y_1 & Y_2 \end{pmatrix}$$

which is equal to

$$\begin{pmatrix} Q_1 A^T + A Q_1 & Y_1^T + B Q_2 \\ Q_2 B^T + Y_1 & Y_2^T + Y_2 \end{pmatrix}.$$

□

However, as this theorem provides a different and original approach, it obviously does not circumvent the input nonnegativity problem as it is only an equivalent representation of the nominal system. Indeed, it is stated that the matrix

$$\begin{pmatrix} Q_1 A^T + A Q_1 & Y_1^T + B Q_2 \\ Q_2 B^T + Y_1 & Y_2^T + Y_2 \end{pmatrix}$$

has to be negative definite in order to positively stabilize the extended system by means of a feedback  $v = \tilde{K}\tilde{x}$ . However, it is known that every principal submatrix of a negative definite matrix is also negative definite (see e.g. (Horn and Johnson 1990, Observation 7.1.2)). This means that  $Q_1 A^T + A Q_1$  has to be negative definite and thus that the initial system should be already stable.

### 1.3 Designing positively stabilizing feedbacks

The previous chapter gave us an important piece of information concerning the positive stabilization problem: one cannot positively stabilize a positive system by means of a nonnegative input. Now, the goal of this chapter is to provide a systematic way to parameterize all positively stabilizing state feedbacks for a positive LTI finite-dimensional system, and we have a first indication of the structure of the feedback matrix. Indeed, as we know that  $u = Kx$  with  $x$  positive, it is clear that  $K$  cannot be a nonnegative matrix as the input  $u$  would also be nonnegative and the system would not be stabilized.

Let us consider the positive system described by the equation  $\dot{x} = Ax + Bu$  where the input  $u = Kx$  is a feedback controller, with  $A \in \mathbb{R}^{n \times n}$  a Metzler matrix,  $B \in \mathbb{R}^{n \times m}$  a nonnegative matrix and  $K \in \mathbb{R}^{m \times n}$  the state feedback matrix. Regarding Proposition 1.1.3, the closed-loop system is positive and stable if

1.  $(A + BK)$  is a Metzler matrix, and
2. there exists a vector  $v \gg 0$  such that  $(A + BK)v \ll 0$ .

These are simple algebraic conditions that one can model and solve in order to find  $K$ . However, the second condition leads to a set of quadratic strict inequalities as both  $v$  and  $K$  are unknown variables. Solving a set of quadratic inequalities is a feasible problem but it is also far more complicated and demanding than solving a linear problem. One way to make the problem easier is to consider single-input systems only. In that case, the structure of the matrix  $BK$  is simpler, though the quadratic property of the problem still holds. To get rid of that issue, let us go further and consider a particular class of systems so that we can provide the reader with a systematic way to design the general expression of any positively stabilizing feedback. More precisely, let us consider single-input positive LTI systems described by the equation  $\dot{x} = Ax + bu$  where  $A \in \mathbb{R}^{n \times n}$  is Metzler, and  $b \in \mathbb{R}^n$  is nonnegative and has only one non-null entry (w.l.o.g., the first one). The particular structure of  $b$  is pretty common, notably when applying the finite difference method to PDE systems with boundary control (see Section 2.2).

The forthcoming theorem is analogous to (Ait Rami 2011, Theorem 3.1), though the latter concerns multi-input single-output systems with output feedback while the result described in this work concerns single-input multi-output systems with state feedback. Let us first introduce the two following concepts, and a lemma that is known as the Farkas-Minkowski-Weyl theorem. The latter and its proof come from (Schrijver 1998, Corollary 7.1a) and will be very useful in the results to come.

**Definition 6** A cone  $C$  is polyhedral if  $C = \{x \mid Ax \leq 0\}$  for some matrix  $A$ .

**Definition 7** The cone (finitely) generated by the vectors  $x_1, \dots, x_n$ , that is denoted by  $\mathbf{c}\{x_1, \dots, x_n\}$ , is the set  $\{\alpha_1 x_1 + \dots + \alpha_n x_n \mid \alpha_1, \dots, \alpha_n \geq 0\}$ , i.e. the smallest convex cone containing  $x_1, \dots, x_n$ . The interior of the cone generated by the vectors  $x_1, \dots, x_n$  will be denoted by  $\mathring{\mathbf{c}}\{x_1, \dots, x_n\}$ .

**Lemma 1.3.1** *A convex cone is polyhedral if and only if it is finitely generated.*

**Proof.** To prove sufficiency, let  $x_1, \dots, x_m$  be vectors in  $\mathbb{R}^n$ . We show that  $\mathfrak{c}\{x_1, \dots, x_m\}$  is polyhedral. We may assume that  $x_1, \dots, x_m$  span  $\mathbb{R}^n$  (see (Schrijver 1998, Corollary 7.1a)). Now consider all linear half-spaces  $H = \{x \mid cx \leq 0\}$  of  $\mathbb{R}^n$  such that  $x_1, \dots, x_m$  belong to  $H$  and such that  $\{x \mid cx = 0\}$  is spanned by  $n - 1$  linearly independent vectors from  $\{x_1, \dots, x_m\}$ . By (Schrijver 1998, Theorem 7.1),  $\mathfrak{c}\{x_1, \dots, x_m\}$  is the intersection of these half-spaces. Since there are only finitely many such half-spaces, the cone is polyhedral. To prove necessity, let  $C$  be a polyhedral cone that is defined by  $C = \{x \mid r_1^T x \leq 0, \dots, r_m^T x \leq 0\}$  for certain column vectors  $r_1, \dots, r_m$ . As each finitely generated cone is polyhedral, there exist column vectors  $s_1, \dots, s_t$  such that

$$\mathfrak{c}\{r_1, \dots, r_m\} = \{x \mid s_1^T x \leq 0, \dots, s_t^T x \leq 0\}. \quad (1.3.1)$$

We show that  $C = \mathfrak{c}\{s_1, \dots, s_t\}$ , implying that  $C$  is finitely generated. Indeed, one can see that  $\mathfrak{c}\{s_1, \dots, s_t\} \subseteq C$ , as  $s_1, \dots, s_t \in C$ , since  $s_j^T r_i \leq 0$  for  $i = 1, \dots, m$  and  $j = 1, \dots, t$ , by (1.3.1). Now, suppose that  $w \notin \mathfrak{c}\{s_1, \dots, s_t\}$  for some  $w \in C$ . As  $\mathfrak{c}\{s_1, \dots, s_t\}$  is polyhedral, there exists a vector  $y$  such that  $y^T s_1, \dots, y^T s_t \leq 0$  and  $y^T w > 0$ . Hence by (1.3.1),  $y \in \mathfrak{c}\{r_1, \dots, r_m\}$ , and hence  $y^T x \leq 0$  for all  $x \in C$ . This contradicts the fact that  $w$  is in  $C$  and  $y^T w > 0$ .  $\square$

We can now state and prove the result concerning the parameterization of all positively stabilizing feedbacks, that we illustrate by a standard example in Section 2.2.

**Theorem 1.3.2** *Let a LTI positive system described by the equation  $\dot{x} = Ax + bu$  where  $A$  is a Metzler matrix and  $b$  is a positive vector that has only its first entry different from zero, i.e.  $b = b_1 e_1$  where  $b_1 > 0$ . Then*

1. *the feedback matrix  $k = (k_1 \quad \dots \quad k_n)$  is positively stabilizing for the system if and only if*

$$k_1 = \frac{-a_{11}v_1 - (a_{12} + b_1 k_2)v_2 - \dots - (a_{1n} + b_1 k_n)v_n - \omega}{b_1 v_1}$$

and

$$k_i \geq \frac{-a_{1i}}{b_1} \quad i = 2, \dots, n$$

where  $\omega > 0$  is a free parameter, and  $v \in \mathbb{R}^n$  is strictly positive and solution of the strict inequalities set

$$\begin{aligned} -a_{21}v_1 - \dots - a_{2n}v_n &> 0, \\ &\vdots \\ -a_{n1}v_1 - \dots - a_{nn}v_n &> 0, \end{aligned} \quad (1.3.2)$$

2. *the set of solutions of (1.3.2) with the strict positivity constraint over  $v$  is given by  $\mathfrak{c}\{s_1, \dots, s_r\}$  where  $r \leq 2n - 1$  and the column vectors  $s_1, \dots, s_r$  are such that  $\mathfrak{c}\{a_2^T, \dots, a_n^T, -e_1, \dots, -e_n\} = \{x \mid s_1^T x \leq 0, \dots, s_r^T x \leq 0\}$  where  $a_i$  denotes the  $i^{\text{th}}$  row of  $A$  and  $e_i$  the  $i^{\text{th}}$  vector of the canonical basis of  $\mathbb{R}^n$ .*

**Proof.** 1. It is straightforward to show that any feedback matrix  $k$  as described above positively stabilizes the system. Now let us consider a general feedback matrix  $k = (k_1 \ \dots \ k_n)$  which yields the closed-loop matrix

$$A + bk = \begin{pmatrix} a_{11} + b_1 k_1 & \dots & a_{1n} + b_1 k_n \\ a_{21} & \dots & a_{2n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}.$$

To ensure that positivity is maintained, this matrix has to be Metzler. This yields the conditions

$$k_i \geq \frac{-a_{1i}}{b_1} \quad i = 2, \dots, n$$

with  $k_1$  free. Regarding the stability property, let us consider Proposition 1.1.3:  $A + bk$  is stable if and only if one can find a vector  $v \gg 0$  such that  $(A + bk)v \ll 0$ . This leads to the following set of strict inequalities

$$\begin{aligned} (a_{11} + b_1 k_1)v_1 + \dots + (a_{1n} + b_1 k_n)v_n &< 0 \\ a_{21}v_1 + \dots + a_{2n}v_n &< 0 \\ &\vdots \\ a_{n1}v_1 + \dots + a_{nn}v_n &< 0 \\ v_1 &> 0 \\ &\vdots \\ v_n &> 0 \end{aligned}$$

that we can rewrite in a more practical way as

$$\begin{aligned} -(a_{11} + b_1 k_1)v_1 - \dots - (a_{1n} + b_1 k_n)v_n &= \omega \\ -a_{21}v_1 - \dots - a_{2n}v_n &> 0 \\ &\vdots \\ -a_{n1}v_1 - \dots - a_{nn}v_n &> 0 \\ v_1 &> 0 \\ &\vdots \\ v_n &> 0 \end{aligned}$$

where  $\omega > 0$  is a free parameter. As only the first equation depends on the entries of  $k$ , we can express

$$k_1 = \frac{-a_{11}v_1 - (a_{12} + b_1 k_2)v_2 - \dots - (a_{1n} + b_1 k_n)v_n - \omega}{b_1 v_1}$$

with  $v$  any solution of the strict inequalities set above.

2. Now we establish the structure of  $v$ . The proof is based on Lemma 1.3.1. Consider the strict inequalities set (1.3.2) with the positivity constraint over the components of  $v$ , both with non-strict inequality signs, that is,

$$\begin{aligned} a_{21}v_1 + \dots + a_{2n}v_n &\leq 0 \\ &\vdots \\ a_{n1}v_1 + \dots + a_{nn}v_n &\leq 0 \\ -v_1 &\leq 0 \\ &\vdots \\ -v_n &\leq 0. \end{aligned}$$

The set of solutions is  $S := \{v \mid a_2v \leq 0, \dots, a_nv \leq 0, -e_1^T v \leq 0, \dots, -e_n^T v \leq 0\}$ . As any finitely generated cone is polyhedral, there exist column vectors  $s_1, \dots, s_r$  such that

$$\mathfrak{c}\{a_2^T, \dots, a_n^T, -e_1, \dots, -e_n\} = \{x \mid s_1^T x \leq 0, \dots, s_r^T x \leq 0\}. \quad (1.3.3)$$

This implies that, for  $j = 1, \dots, r$ ,  $a_i s_j \leq 0$  for  $i = 2, \dots, n$  and  $-e_i^T s_j \leq 0$  for  $i = 1, \dots, n$ . Therefore,  $s_1, \dots, s_r \in S$  and  $\mathfrak{c}\{s_1, \dots, s_r\} \subseteq S$ . Now, let  $w \in S$  and suppose that  $w \notin \mathfrak{c}\{s_1, \dots, s_r\}$ . As  $\mathfrak{c}\{s_1, \dots, s_r\}$  is polyhedral (see the proof of Lemma 1.3.1), there exists a vector  $y$  such that  $y^T s_1 \leq 0, \dots, y^T s_r \leq 0$  and  $y^T w > 0$ . By (1.3.3),  $y \in \mathfrak{c}\{a_2^T, \dots, a_n^T, -e_1, \dots, -e_n\}$  and then  $y^T v \leq 0$  for all  $v \in S$ . This contradicts the facts that  $w \in S$  and  $y^T w > 0$ . As  $\mathfrak{c}\{s_1, \dots, s_r\} = S$  and as strict inequalities are concerned, we take the interior of the cone generated by  $s_1, \dots, s_r$ , which is  $\hat{\mathfrak{c}}\{s_1, \dots, s_r\}$ .  $\square$

## 1.4 Invariant stabilization

In the previous sections we deal with positivity only, i.e. the invariance of the non-negative orthant of the state space. However, one could ask for the invariance of a different subset. This could happen if, for example, there is either a positive or negative tolerance threshold or margin on the state value. Systems involving temperatures or speeds as state variables, among many other cases, fall within this category. The goal of this chapter is to show how one can extend the positivity upholding results to the invariance of a different subset.

### 1.4.1 Cone invariance

Consider the LTI system described by the equation

$$\dot{x} = Ax, \quad (1.4.1)$$

and let us introduce  $\bar{x} \ll 0$  a fixed state in  $\mathbb{R}^n$ . Then we can define the cone

$$C_{\bar{x}} := \{x \in \mathbb{R}^n \mid x_i \geq \bar{x}_i, i = 1, \dots, n\}$$

which is actually a shifted cone as  $C_{\bar{x}} = \mathbb{R}_+^n + \bar{x}$ . Choosing  $\bar{x} = 0$  would bring us back to positivity. Now let us state the following definition from (Beauthier 2011), concerning the invariance of the cone we just defined.

**Definition 8** *The cone  $C_{\bar{x}}$  is said to be invariant with respect to system (1.4.1) if  $C_{\bar{x}}$  is  $e^{At}$ -invariant, i.e.*

$$\forall t \geq 0, \quad e^{At}C_{\bar{x}} \subseteq C_{\bar{x}}$$

or equivalently

$$\forall t \geq 0, \forall x_0 \in C_{\bar{x}}, \quad x(t) := e^{At}x_0 \in C_{\bar{x}}.$$

**Remark 3** *One should note that any element  $y \in \mathbb{R}^n$  in the cone  $C_{\bar{x}}$  can be written as*

$$y = \sum_{j=1}^n \alpha_j \bar{x}_j e_j$$

with  $\alpha_i \leq 1$  for all  $i = 1, \dots, n$ .

As in Section 1.1, one could consider a more geometric perspective and express the invariance of a cone  $C_{\bar{x}}$  in the following way: if the initial state is in the cone  $C_{\bar{x}}$ , the state trajectory remains in the cone at all time (see Figure 1.2).

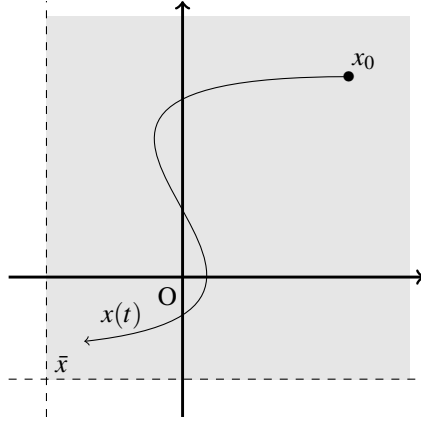


Figure 1.2: The state trajectory remaining in the cone  $C_{\bar{x}}$

For convenience, let us introduce the concept of  $C_{\bar{x}}$ -invariance that will be used in the forthcoming results in order to simplify the notations.

**Definition 9** *System (1.4.1) (its feedback form respectively) is said to be  $C_{\bar{x}}$ -invariant if the cone  $C_{\bar{x}}$  is invariant with respect to system (1.4.1) (its feedback form respectively), i.e. if  $C_{\bar{x}}$  is  $e^{At}$ -invariant ( $e^{(A+BK)t}$ -invariant respectively).*



So the stabilization problem is now concerned with existence conditions and the computation of a feedback matrix  $K$  that leads to the stability of the closed-loop system and to the invariance of the cone  $C_{\bar{x}}$ . In order to solve this problem, let us introduce a result from (Beauthier 2011) that derives from (Castelan and Hennet 1993, Proposition 1), and which provides a very useful piece of information concerning the invariance of  $C_{\bar{x}}$ . Before we state this result, we need the following lemma and its proof which provides an interesting intuitive approach of the cone invariance problem, that also come from (Beauthier 2011).

**Lemma 1.4.1** *If  $C_{\bar{x}}$  is invariant with respect to system (1.4.1), if for all  $t \geq 0$ ,  $x(t) = e^{At}x_0$ , where  $x_0$  is any initial state in  $C_{\bar{x}}$  and if there exists  $i \in \{1, \dots, n\}$  such that  $x_i(t) = \bar{x}_i$ , then  $\dot{x}_i(t) \geq 0$ .*

**Proof.** Suppose that for some  $x_0 \geq \bar{x}$ , there exist at least one index  $i = 1, \dots, n$  and a time  $t \geq 0$  such that  $x_i(t) = \bar{x}_i$  and  $\dot{x}_i(t) < 0$ . Now, by assumption, for all  $x_0 \geq \bar{x}$ ,  $x(t) \geq \bar{x}$  where  $x(t)$  is solution of system (1.4.1). In particular,  $x_i(t) \geq -\bar{x}_i$  for all  $t \geq 0$ . Moreover,  $\dot{x}(t) = Ax(t) = Ae^{At}x_0 = e^{At}Ax_0$  for all  $t \geq 0$ . Since the function  $\dot{x}_i(\cdot)$  is continuous on  $\mathbb{R}_+$ ,  $\dot{x}_i(t) < 0$  implies that there exists  $t_1 > t$  such that for all  $\tau \in [t, t_1]$ ,  $\dot{x}_i(\tau) < 0$ , that means  $x_i(\cdot)$  is strictly decreasing on  $[t, t_1]$  with  $x_i(t) = \bar{x}_i$ . Therefore  $x_i(t_1) < x_i(t) = \bar{x}_i$  for all  $\tau \in [t, t_1]$ . It follows that  $x(t_1) \notin C_{\bar{x}}$ . On the other hand, since  $x(t) \in C_{\bar{x}}$  for all  $t \geq 0$ ,  $x(t_1) = e^{A(t_1-t)}x(t) \in C_{\bar{x}}$ . This contradicts the fact that  $x(t_1) \notin C_{\bar{x}}$ . Thus  $\dot{x}_i(t) \geq 0$ .  $\square$

Now we state and prove the result (see (Beauthier 2011, Theorem 1.1.5)). The proof also provides an interesting approach and a better understanding of the forthcoming results and remarks.

**Theorem 1.4.2** *The cone  $C_{\bar{x}}$  is invariant with respect to system (1.4.1) if and only if*

$$A \text{ is a Metzler matrix} \quad (1.4.2)$$

and

$$A\bar{x} \geq 0. \quad (1.4.3)$$

**Proof.** Since for all  $x_0 \geq \bar{x}$ ,  $x(t) \geq \bar{x}$  for all  $t \geq 0$ , by Lemma 1.4.1,

$$\dot{x}_i(t) = \sum_{j=1}^n a_{ij}x_j(t) \geq 0$$

for all  $i = 1, \dots, n$  such that  $x_i(t) = \bar{x}_i$ . First, let us consider  $x(0) = x_0 := \bar{x} \in C_{\bar{x}}$ . It follows from Lemma 1.4.1 applied in  $t = 0$  that for all  $i = 1, \dots, n$ ,  $\dot{x}_i(0) \geq 0$ , or equivalently  $\dot{x}(0) = A\bar{x} \geq 0$ , hence condition (1.4.3) holds. Now by Remark 3, let us consider

$$x(0) = \sum_{j=1}^n \alpha_j \bar{x}_j e_j$$

where  $\alpha_i = 1$  and  $\alpha_j < 0$  for some arbitrarily fixed  $i, j = 1, \dots, n$  such that  $i \neq j$ , and  $\alpha_k = 0$  for all  $k \neq i, j$ . Then,

$$\dot{x}_i(0) = a_{ii}\bar{x}_i + a_{ij}\alpha_j\bar{x}_j \geq 0$$

or equivalently

$$a_{ij} \geq -\frac{\bar{x}_i}{\bar{x}_j} a_{ii} \frac{1}{\alpha_j}$$

with  $\alpha_j < 0$  for  $j \neq i$ . Letting  $\alpha_j$  tend to  $-\infty$ , it follows that  $a_{ij} \geq 0$ . Since  $i$  and  $j$  were arbitrarily fixed, one can conclude that condition (1.4.2) holds. Now assume that conditions (1.4.2) and (1.4.3) hold, i.e.

$$\forall t \geq 0, e^{At} \geq 0 \quad \text{and} \quad A\bar{x} \geq 0.$$

Then, for all  $t \geq 0$ ,

$$e^{At}\bar{x} = \bar{x} + \int_0^t e^{A\tau} A\bar{x} d\tau,$$

where  $e^{A\tau} A\bar{x} \geq 0$  for all  $\tau \in [0, t]$ . Therefore,  $e^{At}\bar{x} \geq \bar{x}$ . Hence,

$$e^{At}x_0 \geq e^{At}\bar{x} \geq \bar{x}$$

for all  $t \geq 0$  and for every  $x_0 \in \mathbb{R}^n$  such that  $x_0 \geq \bar{x}$ . One can then conclude that  $x(t) \geq \bar{x}$ .  $\square$

This result actually provides implicit information on the matrix  $A$ . Indeed, conditions (1.4.2) and (1.4.3) imply that for all  $i = 1, \dots, n$ ,

$$(A\bar{x})_i = \sum_{j=1}^n a_{ij}\bar{x}_j = a_{ii}\bar{x}_i + \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij}\bar{x}_j \geq 0$$

and then

$$a_{ii} \leq -\sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} \frac{\bar{x}_j}{\bar{x}_i} \leq 0$$

as  $a_{ij} \geq 0$  for all  $j \neq i$ , and  $\bar{x}_i$  and  $\bar{x}_j$  are negative. This leads to the conclusion that  $a_{ii} \leq 0$  for all  $i = 1, \dots, n$ , i.e. that the diagonal elements of the matrix  $A$  should be nonpositive. This is a property one could already deduce from Remark 3, as considering  $x(0) = y$  with  $\alpha_i = 1$  for some  $i \in \{1, \dots, n\}$  and  $\alpha_j = 0$  for all  $j \neq i$ , we obtain  $\dot{x}_i(0) = a_{ii}\bar{x}_i \geq 0$  and then  $a_{ii} \leq 0$  for all  $i = 1, \dots, n$ . Moreover, conditions (1.4.2) and (1.4.3) can be seen as a weighted diagonal dominance condition on the matrix  $A$ . Indeed, condition (1.4.2) implies that the off-diagonal elements of  $A$  are nonnegative, and condition (1.4.3) implies that the matrix product of  $A$  by a strictly negative vector should be nonnegative. Then

$$|a_{ii}| \geq \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} \frac{\bar{x}_j}{\bar{x}_i} \geq 0$$

for all  $i = 1, \dots, n$ .

Now, regarding Section 1.3, one can easily deduce that the closed-loop system is  $C_{\bar{x}}$ -invariant and stable if

1.  $(A + BK)$  is a Metzler matrix,
2.  $(A + BK)\bar{x} \geq 0$ , and
3. there exists a vector  $v \gg 0$  such that  $(A + BK)v \ll 0$ .

There are two possible cases. First, the nominal system described by  $\dot{x} = Ax$  is not  $C_{\bar{x}}$ -invariant, i.e. there are no physical constraints to the system. One could ask the  $C_{\bar{x}}$ -invariance of the closed-loop system due to other kinds of restrictions (e.g. technical or economical constraints). In that case, they might have to act on the whole domain to ensure that all algebraic conditions are verified for the closed-loop system. In the other case, the nominal system described by  $\dot{x} = Ax$  is already  $C_{\bar{x}}$ -invariant. Then the problem consists in stabilizing the system while preserving the  $C_{\bar{x}}$ -invariance. To adopt a similar approach as in Section 1.3, let us consider single-input positive LTI systems described by the equation  $\dot{x} = Ax + bu$  where  $A$  is a Metzler matrix such that  $A\bar{x} \geq 0$ , and  $b$  is nonnegative and has only one non-null entry (w.l.o.g., the first one). Some of the algebraic conditions are then already satisfied, and we can state the following result which can easily be deduced from Theorem 1.3.2.

**Theorem 1.4.3** *Let us consider a  $C_{\bar{x}}$ -invariant LTI system described by the equation  $\dot{x} = Ax + bu$  where  $A$  is Metzler such that  $A\bar{x} \geq 0$ , and  $b$  is nonnegative and has only its first entry different from zero. Then the feedback matrix  $k = (k_1 \dots k_n)$  is  $C_{\bar{x}}$ -invariantly stabilizing for the system if and only if*

$$k_1 = \frac{-a_{11}v_1 - (a_{12} + b_1k_2)v_2 - \dots - (a_{1n} + b_1k_n)v_n - \omega}{b_1v_1}$$

such that

$$k_1 \leq \frac{-a_{11}\bar{x}_1 - (a_{12} + b_1k_2)\bar{x}_2 - \dots - (a_{1n} + b_1k_n)\bar{x}_n}{b_1\bar{x}_1}$$

and

$$k_i \geq \frac{-a_{1i}}{b_1} \quad i = 2, \dots, n$$

where  $\omega > 0$  is a free parameter, and  $v \in \mathbb{R}^n$  is strictly positive and solution of the strict inequalities set (1.3.2), the set of solutions of which can be computed in the same way as in Theorem 1.3.2.

### 1.4.2 Sector invariance

In this subsection, we introduce an additional constraint. Let us consider  $\bar{x} \ll 0$  and  $\tilde{x} \gg 0$ , two fixed states in  $\mathbb{R}^n$ , and let us define the cone

$$C_{\bar{x}} := \{x \in \mathbb{R}^n \mid x_i \geq \bar{x}_i, i = 1, \dots, n\} = \mathbb{R}_+^n + \bar{x}$$

as in the previous subsection, and a second cone

$$C_{\tilde{x}}^- := \{x \in \mathbb{R}^n \mid x_i \leq \tilde{x}_i, i = 1, \dots, n\}$$

which is also a shifted cone as  $C_{\tilde{x}}^- = \mathbb{R}_-^n + \tilde{x}$ . We can then define

$$S_{\bar{x}, \tilde{x}} := C_{\bar{x}} \cap C_{\tilde{x}}^- = \{x \in \mathbb{R}^n \mid \bar{x}_i \leq x_i \leq \tilde{x}_i, i = 1, \dots, n\}$$

which is a subset we will refer to as a sector, as the state variables now have both lower and upper bounds. The concepts of sector and polyhedral invariant sets can be found in (Blanchini and Miani 2007, Sections 3.3 and 4.5) and are closely related to this subsection contents, though in this work we develop our own way to solve the invariance problem. The objective is similar to those of the positivity and cone invariance problems: designing a feedback matrix  $K$  such that the closed-loop system is stable and the state trajectories stay in the sector at all time. This leads to the following definition.

**Definition 10** *The sector  $S_{\bar{x}, \tilde{x}}$  is said to be invariant with respect to system (1.4.1) if  $S_{\bar{x}, \tilde{x}}$  is  $e^{At}$ -invariant, i.e.*

$$\forall t \geq 0, \quad e^{At} S_{\bar{x}, \tilde{x}} \subseteq S_{\bar{x}, \tilde{x}}$$

or equivalently

$$\forall t \geq 0, \forall x_0 \in S_{\bar{x}, \tilde{x}}, \quad x(t) := e^{At} x_0 \in S_{\bar{x}, \tilde{x}}.$$

Ensuring the invariance of a sector could be seen as a Lyapunov stabilization problem, as we force the state to verify some given lower and upper bounds at all time.

**Remark 4** *Regarding Remark 3, one can easily see that any element  $y \in \mathbb{R}^n$  in the sector  $S_{\bar{x}, \tilde{x}}$  can be written as*

$$y = \sum_{j=1}^n \alpha_j \bar{x}_j e_j$$

with  $\frac{\bar{x}_i}{\bar{x}_i} \leq \alpha_i \leq 1$  for all  $i = 1, \dots, n$ , or as

$$y = \sum_{j=1}^n \beta_j \tilde{x}_j e_j$$

with  $\frac{\tilde{x}_i}{\tilde{x}_i} \leq \beta_i \leq 1$  for all  $i = 1, \dots, n$ .

Obviously, one can express the invariance of a sector  $S_{\bar{x}, \tilde{x}}$  in the following way: if the initial state is in the sector  $S_{\bar{x}, \tilde{x}}$ , the state trajectory remains in the sector at all time (see Figure 1.3). Note that one could consider upper bounds on certain state components only, and the upcoming results would require little or no change at all. For example, in Theorem 1.4.4 and Corollaries 1.4.5 and 1.4.6, one would only need to consider the inequalities corresponding to the fixed (finite) components  $\tilde{x}_i$ .

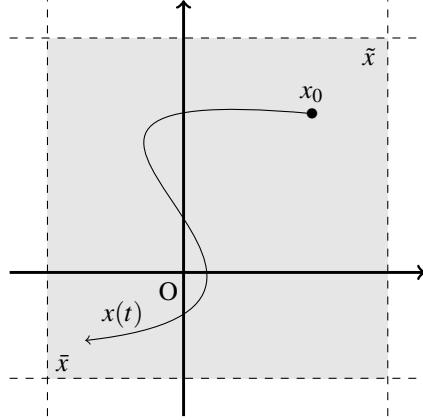


Figure 1.3: The state trajectory remaining in the sector  $S_{\bar{x}, \tilde{x}}$

For convenience, let us introduce the concept of  $S_{\bar{x}, \tilde{x}}$ -invariance in the same way as we did in Subsection 1.4.1 for the  $C_{\bar{x}}$ -invariance.

**Definition 11** System (1.4.1) (its feedback form respectively) is said to be  $S_{\bar{x}, \tilde{x}}$ -invariant if the sector  $S_{\bar{x}, \tilde{x}}$  is invariant with respect to system (1.4.1) (its feedback form respectively), i.e. if  $S_{\bar{x}, \tilde{x}}$  is  $e^{At}$ -invariant ( $e^{(A+BK)t}$ -invariant respectively).

In order to ensure the invariance of the sector, the state components derivatives need to be nonnegative (respectively nonpositive) when a lower bound (respectively an upper bound) is reached, i.e. for all  $i = 1, \dots, n$ ,

$$\begin{cases} \dot{x}_i \geq 0 & \text{if } x_i = \bar{x}_i, \\ \dot{x}_i \leq 0 & \text{if } x_i = \tilde{x}_i. \end{cases}$$

or equivalently

$$\begin{cases} (A + BK)_i x \geq 0 & \text{if } x_i = \bar{x}_i, \\ (A + BK)_i x \leq 0 & \text{if } x_i = \tilde{x}_i. \end{cases}$$

where  $(A + BK)_i$  denotes the  $i^{\text{th}}$  row of  $(A + BK)$ . This leads to the following result.

**Theorem 1.4.4** *The sector  $S_{\bar{x}, \bar{x}}$  is invariant with respect to system (1.4.1) if and only if*

$$\begin{aligned} a_{ii}\bar{x}_i + \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij}\alpha_j\bar{x}_j &\geq 0 \\ a_{ii}\tilde{x}_i + \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij}\alpha_j\bar{x}_j &\leq 0 \end{aligned} \quad (1.4.4)$$

for all  $i, j \in \{1, \dots, n\}$ , with  $\alpha_j \in [\tilde{x}_j/\bar{x}_j, 1]$ .

**Proof.** First recall that, by Remark 4,  $x \in S_{\bar{x}, \bar{x}}$  if and only if  $x_i = \alpha_i \bar{x}_i$ , for all  $i = 1, \dots, n$ , where  $\frac{\tilde{x}_i}{\bar{x}_i} \leq \alpha_i \leq 1$ . Suppose that  $S_{\bar{x}, \bar{x}}$  is  $e^{At}$ -invariant. If  $x_i(t) = \bar{x}_i$  for some  $t > 0$ , then

$$\dot{x}_i(t) = (Ax)_i = a_{ii}\bar{x}_i + \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij}\alpha_j\bar{x}_j \geq 0$$

and if  $x_i(t) = \tilde{x}_i$  for some  $t > 0$ , then

$$\dot{x}_i(t) = (Ax)_i = a_{ii}\tilde{x}_i + \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij}\alpha_j\bar{x}_j \leq 0.$$

Now suppose condition (1.4.4) holds: as it is valid for any  $\alpha_j \in [\tilde{x}_j/\bar{x}_j, 1]$ , the sufficiency is trivially obtained.  $\square$

Although this result provides a necessary and sufficient condition for the  $e^{At}$ -invariance of the sector  $S_{\bar{x}, \bar{x}}$ , it is of little use in an application context as one would have to consider an infinite uncountable number of different cases in order to design the feedback matrix  $K$  for the closed-loop system. We then introduce the following corollary.

**Corollary 1.4.5** *The sector  $S_{\bar{x}, \bar{x}}$  is invariant with respect to system (1.4.1) if and only if*

$$\begin{aligned} a_{ii}\bar{x}_i + \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij}\alpha_j\bar{x}_j &\geq 0 \\ a_{ii}\tilde{x}_i + \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij}\alpha_j\bar{x}_j &\leq 0 \end{aligned} \quad (1.4.5)$$

for all  $i, j \in \{1, \dots, n\}$ , with  $\alpha_j \in \{\tilde{x}_j/\bar{x}_j, 1\}$ .

**Proof.** The necessity is obvious as it is a particular case of the conditions of Theorem 1.4.4. Now suppose condition (1.4.5) holds and  $S_{\bar{x}, \bar{x}}$  is not  $e^{At}$ -invariant, i.e. there exist  $t > 0$  and  $i \in \{1, \dots, n\}$  such that (w.l.o.g.)  $x_i(t) = \bar{x}_i$  and  $\dot{x}_i(t) < 0$ . Then

$$\begin{aligned} \dot{x}_i(t) &= a_{ii}\bar{x}_i + \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij}\alpha_j\bar{x}_j \\ &= a_{ii}\bar{x}_i + \sum_{j \in S_+} a_{ij}\alpha_j\bar{x}_j + \sum_{k \in S_-} a_{ik}\alpha_k\bar{x}_k \\ &< 0 \end{aligned}$$

where

$$S_+ = \{j \in \{1, \dots, n\} \mid j \neq i, a_{ij} \geq 0\}$$

and

$$S_- = \{k \in \{1, \dots, n\} \mid k \neq i, a_{ik} < 0\}.$$

Now, as

$$\sum_{j \in S_+} a_{ij}\alpha_j\bar{x}_j \geq \sum_{j \in S_+} a_{ij}\bar{x}_j$$

and

$$\sum_{k \in S_-} a_{ik}\alpha_k\bar{x}_k \geq \sum_{k \in S_-} a_{ik} \frac{\tilde{x}_k}{\bar{x}_k} \bar{x}_k = \sum_{k \in S_-} a_{ik}\tilde{x}_k$$

we obtain

$$0 \leq a_{ii}\bar{x}_i + \sum_{j \in S_+} a_{ij}\bar{x}_j + \sum_{k \in S_-} a_{ik}\tilde{x}_k \leq a_{ii}\bar{x}_i + \sum_{j \in S_+} a_{ij}\alpha_j\bar{x}_j + \sum_{k \in S_-} a_{ik}\alpha_k\bar{x}_k < 0$$

which is a contradiction.  $\square$

In Corollary 1.4.5, the condition is refined: it is still necessary and sufficient, but there is a finite number of conditions. It is then possible to implement them and try to solve the problem in order to design a  $S_{\bar{x}, \bar{x}}$ -invariantly stabilizing feedback. Finally, we introduce a second corollary.

**Corollary 1.4.6** *The sector  $S_{\bar{x}, \bar{x}}$  is invariant with respect to system (1.4.1) if and only if*

$$\begin{aligned} a_{ii}\bar{x}_i + \sum_{j \in S_+} a_{ij}\bar{x}_j + \sum_{k \in S_-} a_{ik}\tilde{x}_k &\geq 0 \\ a_{ii}\tilde{x}_i + \sum_{j \in S_-} a_{ij}\bar{x}_j + \sum_{k \in S_+} a_{ik}\tilde{x}_k &\leq 0 \end{aligned} \tag{1.4.6}$$

for all  $i, j \in \{1, \dots, n\}$ , where

$$S_+ = \{j \in \{1, \dots, n\} \mid j \neq i, a_{ij} \geq 0\}$$

and

$$S_- = \{j \in \{1, \dots, n\} \mid j \neq i, a_{ij} < 0\}.$$

**Proof.** The necessity is obvious as it is a particular case of Theorem 1.4.4 and Corollary 1.4.5. The sufficiency is obvious regarding the proof of the sufficiency of Corollary 1.4.5.  $\square$

Corollary 1.4.6 provides an even more refined condition as it actually deals with the few "worst" cases. However, the conditions imply that we need to know the sign of the off-diagonal entries of the matrix  $A$ , making it obsolete if we wish to use the conditions in order to design a feedback matrix for the closed-loop system.

Now let us focus on the stabilization part. One should note that, as the Metzler property is not required anymore, it is not possible to use Proposition 1.1.3 as before. We then have to consider another result to ensure the stability of the closed-loop system. One possible and standard way to do so is to use a Lyapunov inequality as we did in Subsection 1.2.2 (see e.g. (Boyd et al. 1994)). The result is as follows.

**Theorem 1.4.7** *Let a LTI system be described by the equation  $\dot{x} = Ax + Bu$ . A feedback matrix  $K$  is stabilizing for the system if and only if there exists a positive definite matrix  $Q$  such that  $Q(A + BK)^T + (A + BK)Q$  is negative definite.*

As both  $Q$  and  $K$  are unknowns, this problem is nonlinear. One could then use the change of variables  $Y = KQ$  so that the problem becomes finding  $Q$  positive definite and  $Y$  such that  $QA^T + Y^T B^T + AQ + BY$  is negative definite. The problem is then linear, and one can compute the feedback matrix afterwards by  $K = YQ^{-1}$  ( $Q$  is always invertible as it is positive definite). However, the change of variables makes the  $S_{\bar{x}, \bar{x}}$ -invariance conditions nonlinear. It is still possible to gather the two parts and to state the following theorem which is a direct consequence of the previous results.

**Theorem 1.4.8** *Let a LTI system described by the equation  $\dot{x} = Ax + Bu$ , and let  $X \in \mathbb{R}^{n \times 2^n}$  the matrix which columns are all the possible vectors such that  $X_{ij} = \bar{x}_i$  or  $X_{ij} = \bar{x}_i$ , with  $j = 1, \dots, 2^n$ . A feedback matrix  $K$  is  $S_{\bar{x}, \bar{x}}$ -invariantly stabilizing for the system if and only if there exist a positive definite matrix  $Q$ , a matrix  $Y$  and a parameter matrix  $W \in \mathbb{R}^{n \times 2^n}$  such that*

1.  $QA^T + Y^T B^T + AQ + BY$  is negative definite,
2.  $(AQ + BY)_{i, \cdot} W_{\cdot, j} \begin{cases} \geq 0 & \text{if } w_j \leq 0 \\ \leq 0 & \text{if } w_j \geq 0 \end{cases}$
3.  $QW_{\cdot, j} = X_{\cdot, j}$ ,
4.  $\text{sgn}(W_{i, j}) = \text{sgn}(X_{i, j})$ ,

for all  $i = 1, \dots, n$ , for all  $j = 1, \dots, 2^n$ , where  $A_{i, \cdot}$  and  $A_{\cdot, j}$  denote the  $i^{\text{th}}$  row and the  $j^{\text{th}}$  column of the matrix  $A$ , respectively.

However, this method is numerically very costly, as illustrated in Subsection 3.2.3, and should be applied to low-dimensional systems only.



### 1.4.3 Lyapunov level sets

In Subsections 1.4.1 and 1.4.2, the feedback matrices were designed to ensure both the invariance and the stability at the same time for the closed-loop system, for all initial conditions in the cone or sector. Depending on the nominal system and the discretized model, this condition could appear to be too strong (as illustrated in Section 3.2) and one could wish to weaken it. In this section, it is shown how to design a stabilizing feedback and then to compute a set of initial conditions that ensure that the state remains in a given cone or sector.

Consider a LTI stable system described by the equation  $\dot{x} = Ax$ . As it is stable, we know that there exists a positive definite symmetric matrix  $P$  such that  $A^T P + PA$  is a negative definite matrix (see e.g. (Haddad et al. 2010, Theorem 2.12)). Let us define the Lyapunov function  $V$  such that  $V(x) := x^T P x$ . Obviously,  $V(x) \geq 0$  for all  $x$ , and

$$\begin{aligned} \dot{V}(x) &= \dot{x}^T P x + x^T P \dot{x} \\ &= (Ax)^T P x + x^T P Ax \\ &= x^T (A^T P + PA) x \\ &< 0. \end{aligned}$$

We can then introduce the Lyapunov level sets which are the subsets defined by

$$V(x) = x^T P x = r$$

with  $r \in \mathbb{R}_+$ . Considering the expression of  $V(x)$ , and supposing that  $P$  is fixed, one can easily see that these subsets are actually ellipsoids that share the origin as their center. The concept of ellipsoidal invariant sets with Lyapunov functions has been approached in (Blanchini and Miani 2007, Section 4.4), though not in the same way nor for the same objective. Now by definition of the Lyapunov function, we know that for a stable system the state trajectory (strictly) decreases along the Lyapunov function level sets. This means that if we start with an initial condition that is on or inside a given level set, the state trajectory will remain inside it. If we consider  $\bar{x} \ll 0$  a fixed state in  $\mathbb{R}^n$ , and the cone

$$C_{\bar{x}} := \{x \in \mathbb{R}^n \mid x_i \geq \bar{x}_i, i = 1, \dots, n\}$$

as in Subsection 1.4.1, the problem becomes finding the largest  $r \in \mathbb{R}_+$  such that the corresponding Lyapunov level set is included in the cone, i.e. finding the largest  $r \in \mathbb{R}_+$  such that  $V(x_0) = r$  implies that  $x(t) \in C_{\bar{x}}$  for all  $t > 0$ . Obviously, such a  $r$  should lie between zero and  $V(\bar{x})$ . Moreover, by geometric intuition one can deduce that the level set corresponding to the largest  $r$  should be tangent to one of the  $n$  hyperplanes defining the cone (see Figure 1.4).

**Remark 5** In (Boyd et al. 1994, Section 3.7) it is explained how to approximate a subset by an ellipsoid, notably how to find the ellipsoid of smallest volume that contains a polytope described by a set of linear inequalities. Though we actually want

the ellipsoid to be inside the polytope, this could be a starting point for a new design method.

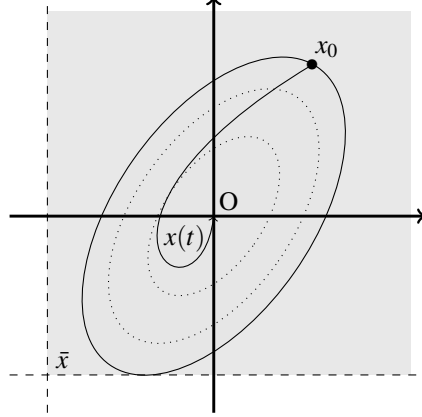


Figure 1.4: The state trajectory remaining in the level set

One way to bring a solution to the problem is to choose

$$r := \min_{i \in \{1, \dots, n\}} \{r_i\}$$

where  $r_i$  is such that the ellipsoid corresponding to  $V(x) = r_i$  is tangent to the  $i^{\text{th}}$  hyperplane. However, solving  $n$  systems of equations of the type

$$\begin{cases} x^T P x &= r_i \\ x_i &= \bar{x}_i \end{cases} \quad (1.4.7)$$

with  $i \in \{1, \dots, n\}$  fixed would be very costly. To make the problem much simpler, one could recall that the intersection of an ellipsoid and an hyperplane is still an ellipsoid (of smaller dimension). We could then find a condition so that the intersection is degenerate into a single point. In order to do so, let us state the following lemma (see e.g. (Lasley 1957)).

**Lemma 1.4.9** *An ellipsoid defined by the equation*

$$ax^2 + 2bxy + cy^2 + 2dx + 2ey + f = 0$$

*is degenerate if and only if its discriminant*

$$\Delta = \det \begin{pmatrix} a & b & d \\ b & c & e \\ d & e & f \end{pmatrix}$$

*is equal to zero.*

This result can be generalized to  $n$ -dimensional ellipsoids, and leads to the following theorem.

**Theorem 1.4.10** *Let a LTI stable system be described by the equation  $\dot{x} = Ax$  and let the cone  $C_{\bar{x}} := \{x \in \mathbb{R}^n \mid x_i \geq \bar{x}_i, i = 1, \dots, n\}$  where  $\bar{x} \ll 0$  is a fixed state in  $\mathbb{R}^n$ . The state trajectory  $x(t) = e^{At}x_0$  remains in  $C_{\bar{x}}$  at all time if the initial state  $x_0$  is such that  $x_0^T P x_0 \leq r$  where  $P$  is a positive definite symmetric matrix such that  $A^T P + PA$  is negative definite and*

$$r = \min_{i \in \{1, \dots, n\}} \left\{ \frac{\det(P) \bar{x}_i^2}{\det(P^{[i]})} \right\}$$

where  $P^{[i]}$  is the matrix  $P$  with the  $i^{\text{th}}$  column and the  $i^{\text{th}}$  row removed.

**Proof.** Regarding system (1.4.7) which provides the equation of the  $i^{\text{th}}$  subellipsoid (i.e. the intersection between the level set and the  $i^{\text{th}}$  hyperplane), one can see that the discriminant of the latter is the determinant of the matrix

$$Q = \left( \begin{array}{cccccc|c} P_{1,1} & \dots & P_{1,i-1} & P_{1,i+1} & \dots & P_{1,n} & P_{1,i} \bar{x}_i \\ \vdots & & \vdots & \vdots & & \vdots & \vdots \\ P_{i-1,1} & \dots & P_{i-1,i-1} & P_{i-1,i+1} & \dots & P_{i-1,n} & P_{i-1,i} \bar{x}_i \\ P_{i+1,1} & \dots & P_{i+1,i-1} & P_{i+1,i+1} & \dots & P_{i+1,n} & P_{i+1,i} \bar{x}_i \\ \vdots & & \vdots & \vdots & & \vdots & \vdots \\ P_{n,1} & \dots & P_{n,i-1} & P_{n,i+1} & \dots & P_{n,n} & P_{n,i} \bar{x}_i \\ \hline P_{i,1} \bar{x}_i & \dots & P_{i,i-1} \bar{x}_i & P_{i,i+1} \bar{x}_i & \dots & P_{i,n} \bar{x}_i & P_{i,i} \bar{x}_i^2 - r_i \end{array} \right)$$

that we can rewrite as

$$Q = \begin{pmatrix} P^{[i]} & P_{j \neq i, i} \bar{x}_i \\ P_{i, j \neq i} \bar{x}_i & P_{ii} \bar{x}_i^2 - r \end{pmatrix}$$

where we recall that  $P^{[i]}$  is the matrix  $P$  with the  $i^{\text{th}}$  column and the  $i^{\text{th}}$  row removed, and  $P_{j \neq i, i}$  and  $P_{i, j \neq i}$  are respectively the  $i^{\text{th}}$  column and the  $i^{\text{th}}$  row of  $P$  without their  $i^{\text{th}}$  component. By the generalization of Lemma 1.4.9, to ensure that the  $i^{\text{th}}$  subellipsoid is degenerate and thus that the level set is tangent to the  $i^{\text{th}}$  hyperplane, the condition

$$\det(Q) = 0 \tag{1.4.8}$$

has to be verified. Computing the determinant of  $Q$  by the cofactor (or Laplace) expansion, one can see that condition (1.4.8) is verified if and only if

$$r_i = \frac{\det(P) \bar{x}_i^2}{\det(P^{[i]})}$$

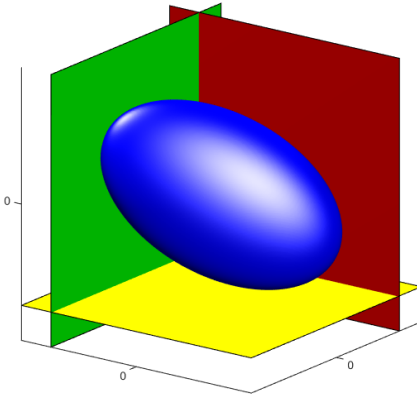
and as we want the ellipsoid

1. to be tangent to (at least) one hyperplane, and
2. not to be secant with any other hyperplane,

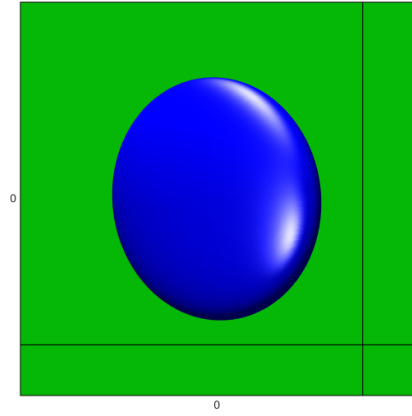
we choose

$$r = \min_{i \in \{1, \dots, n\}} \left\{ \frac{\det(P) \bar{x}_i^2}{\det(P^{[i]})} \right\}.$$

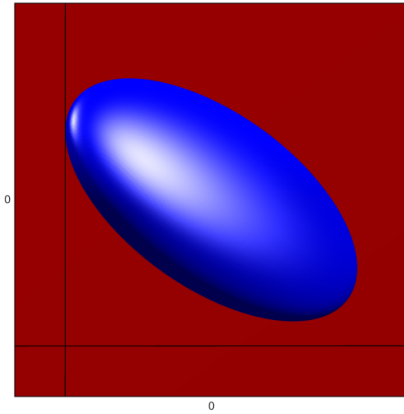
□



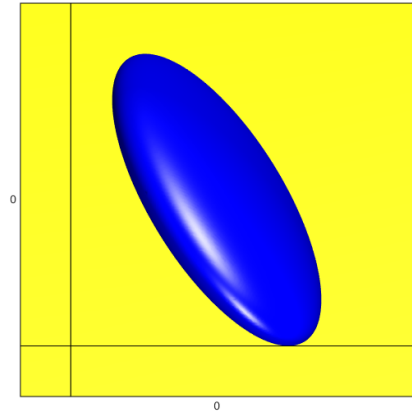
(a) Maximal 3D Lyapunov level set  
(XYZ view)



(b) Maximal 3D Lyapunov level set  
(XZ view)



(c) Maximal 3D Lyapunov level set  
(YZ view)



(d) Maximal 3D Lyapunov level set  
(XY view)

Figure 1.5: Maximal Lyapunov level set (3D example)

This result allows a fast and easy computation of the maximal Lyapunov level set. Figures 1.5a, 1.5b, 1.5c and 1.5d illustrate the result via a simple three-dimensional example.

Note that the algorithm is numerically efficient as the number of calculations increases linearly with the system dimension. The determinant computation will obviously be more costly as the dimension increases, but the algorithm is still very efficient and low cost even for higher values of  $n$ . Also note that  $P$  could be used as a parameter to tune the shape of the level sets, so that they fit the cone or sector better. Thereby, one could hope to improve the design of the initial conditions set.

## Chapter 2

# Positive linear time-invariant PDE systems

### 2.1 Positive stabilization of a positive system by non-negative input

In Section 1.2 we showed the issue of considering a nonnegative input when trying to positively stabilize a finite-dimensional unstable system, and the conclusion was pretty much unequivocal. This property naturally extends to infinite-dimensional unstable systems. In (Achhab and Winkin 2014), it is stated that in order to guarantee that the closed-loop dynamics are nonnegative, the control law is designed such that the resulting input trajectory is nonnegative. As expected, the stabilization problem is then unfeasible. This has been corrected in (Achhab and Winkin 2017) where the input is allowed to be negative but has a lower bound, such that the resulting input trajectory remains in an affine cone.

In order to prove the extension of the result to infinite-dimensional systems, one would naturally turn to a generalization of the Perron-Frobenius theorem (see Theorem 1.2.1). This generalization is known as the Krein-Rutman theorem (see e.g. (Krein and Rutman 1948)) and is stated below.

**Theorem 2.1.1** *Let  $X$  be a Banach space,  $K \subset X$  a solid cone (i.e.  $\mathring{K} \neq \emptyset$ ),  $S : X \rightarrow X$  a compact linear operator which is strongly positive (i.e.  $S(K_0) \subseteq \mathring{K}$ ). Then*

1.  $r(S) > 0$ , and  $r(S)$  is a simple eigenvalue with an eigenvector  $v \in K_0$ , and there is no other eigenvalue with a positive eigenvector;
2.  $|\lambda| < r(S)$  for all eigenvalues  $\lambda \neq r(S)$ .

However, as we are mainly working on  $L^2(a, b)$ , the interior of the cone is actually empty and we cannot apply the theorem. Moreover, we need the result for the generator of a positive  $C_0$ -semigroup, and such an operator is usually not bounded and

certainly not compact. Some additional hypotheses are then needed in order to prove the extension of Theorem 1.2.2, which is stated below.

**Theorem 2.1.2** *Let the positive linear system described by the equation  $\dot{x} = Ax + bu$ , where  $A$  is a Riesz-spectral operator and the generator of a positive  $C_0$ -semigroup  $(T(t))_{t \geq 0}$  on  $L^2(z_1, z_2)$ , and  $b$  is a nonnegative vector. Assume that the spectral bound  $s(A) > -\infty$  is an eigenvalue associated with a positive eigenvector. The system is (exponentially) positively stabilizable by a state feedback  $u = Kx$  such that  $u \in \mathbb{R}_+$  if and only if it is already (exponentially) stable.*

**Proof.** The sufficiency of the condition is trivial as it suffices to take  $K = 0$ , hence  $u = 0$ . Now let us prove the necessity. Suppose that the system is unstable, then the growth bound  $\omega_0(T)$  is positive. As  $s(A) > -\infty$ , it is such that  $s(A) \in \sigma(A)$  (see e.g. (Banasiak and Arlotti 2006, Theorem 3.34)) and it is equal to the growth bound  $\omega_0(T)$  (see e.g. (Banasiak and Arlotti 2006, Theorem 3.37)). Then  $s(A)$  is the dominant eigenvalue  $\lambda_1$  of the adjoint operator  $A^*$ , and by assumption there exists a positive eigenfunction  $\psi_1$  such that  $A^* \psi_1 = \lambda_1 \psi_1$ . Now let us express the state trajectory in the eigenfunctions basis, that is

$$T(t)x_0 = \sum_{i=1}^{\infty} e^{\lambda_i t} \langle \psi_i, x_0 \rangle \varphi_i$$

for all  $x_0$  in the state space. As the eigenfunctions bases are biorthonormal, we know that  $\langle \psi_i, \varphi_j \rangle = \delta_{ij}$  for all  $i, j \in \mathbb{N}$ , and then we can define  $\rho = \langle \psi_1, x \rangle$  and focus on the unstable part of the system relative to  $\lambda_1$ . We then have

$$\begin{aligned} \dot{\rho} &= \langle \psi_1, \dot{x} \rangle \\ &= \langle \psi_1, Ax \rangle + \langle \psi_1, bu \rangle \\ &= \langle A^* \psi_1, x \rangle + \langle \psi_1, bu \rangle \\ &= \lambda_1 \rho + \langle \psi_1, b \rangle u \end{aligned}$$

where  $\langle \psi_1, b \rangle \geq 0$ . If the system was positively stabilizable by a state feedback  $u = Kx$  such that  $u \in \mathbb{R}_+$ , then

$$\rho(t) = e^{\lambda_1 t} \rho_0 + \int_0^t e^{\lambda_1(t-\tau)} \langle \psi_1, b \rangle u(\tau) d\tau$$

would not tend to zero as  $t$  tends to infinity, since  $\lambda_1$ ,  $e^{\lambda_1 t}$ ,  $\rho_0$ ,  $\langle \psi_1, b \rangle$  and  $u$  are all nonnegative, and  $b$  is bounded, thus showing that the system cannot be positively stabilized in this way.  $\square$

However, one can still positively stabilize an unstable system with a nonnegative input, in a certain sense. In (Emirsjlow and Townley 2000) it is explained how one can rewrite a system with boundary control in the boundary conditions as an equivalent model with homogeneous boundary conditions and where the boundary control acts in the dynamics. Two standard examples are given: the Heat equation, which is studied

in Section 2.2, and the Wave equation. The whole process is a bit long and technical, even for standard examples, and is therefore not detailed in this work. We highly encourage the reader to check (Emirsjlow and Townley 2000) for a deeper understanding of the process. Let us simply notify that it includes the computation of two new state spaces, one of them being such that the control operator is bounded (see Subsection 2.3.3), as well as the use of an orthonormal basis of eigenfunctions, the main idea of the process being to split up the solution of the nominal system (where the input acts in the boundary conditions) into a sum of two terms, one of which is a particular solution corresponding to a zero initial condition, that will absorb the control.

Now, as one might expect, it appears the input changes when considering the boundary control in the dynamics, which means a nonnegative boundary control in the boundary conditions could become a nonpositive control in the dynamics. So, a nonnegative input could actually lead to the positive stabilization of an unstable system, depending on which model we are working with. Even more: in some cases, the nonnegativity of the input (acting in the boundary conditions) is actually needed as there is only a change of sign (with intervention of a Dirac delta distribution) when considering the input in the dynamics. This is illustrated in Section 2.2.

## 2.2 Application: the pure diffusion system

### 2.2.1 Modeling

Consider a standard example of unstable positive distributed parameters system: the pure diffusion system described by the partial differential equation

$$\frac{\partial x}{\partial t} = D_a \frac{\partial^2 x}{\partial z^2} \quad (2.2.1)$$

with Neumann boundary conditions

$$\begin{cases} \frac{\partial x}{\partial z}(0, t) = v(t) \\ \frac{\partial x}{\partial z}(L, t) = 0 \end{cases} \quad (2.2.2)$$

where  $v$  is the input,  $D_a$  is the diffusion parameter and  $L$  is the domain length. Note that  $v$  representing the variation of the state at  $z = 0$ , it has no sign restriction. By (Emirsjlow and Townley 2000, Example 2.1) this boundary control system is equivalent to the system described by the PDE

$$\frac{\partial x}{\partial t} = D_a \frac{\partial^2 x}{\partial z^2} + \delta_0 u(t) \quad (2.2.3)$$

with the Dirac delta distribution as control operator and with homogeneous Neumann boundary conditions

$$\begin{cases} \frac{\partial x}{\partial z}(0, t) = 0 \\ \frac{\partial x}{\partial z}(L, t) = 0 \end{cases} \quad (2.2.4)$$



where the new input  $u$  is such that  $u(t) = -v(t)$  for all  $t \geq 0$ . This is actually intuitive as the control acting in the boundary conditions (2.2.2) represents a heat flux going through the surface of the domain entrance, which is reduced to a single point as we are working with a unidimensional model (see e.g. (Vande Wouwer, Saucez and Fernández 2014, Subsection 6.1.3)). If  $v(t)$  is positive, for example, it means that the state value (e.g. the heat) is higher inside the domain than outside it, thus resulting in a negative input  $u(t)$  in the dynamics.

**Remark 6** When considering the PDE (2.2.1) with boundary conditions (2.2.2) and feedback boundary control  $v(t) = \kappa x(0, t)$  with  $\kappa \in \mathbb{R}_{0,+}$ , the input  $v$  is nonnegative. However, when switching to the abstract state equation (2.2.3)-(2.2.4) with input operator as shown in (Emirsjlow and Townley 2000), the sign of the input changes, as  $u(t) = -v(t)$ . The input becomes nonpositive, thus avoiding the issue alluded to in Section 1.2, as explained in the previous section.

**Remark 7** System (2.2.3)-(2.2.4) is a Riesz-spectral system that actually verifies the hypotheses of Theorem 2.1.2 as long as we consider an approximate Dirac delta distribution as the control operator, or if we work in an extended space as in Subsection 2.3.3.

Let us discretize system (2.2.3)-(2.2.4) by the finite difference method, considering  $n$  discretization points  $z_i$ ,  $i = 1, \dots, n$ , with  $z_1 = 0$ ,  $z_n = L$  and  $\Delta z = L/(n-1)$  the discretization step. At  $z_1$  we obtain

$$\begin{aligned} \dot{x}(z_1, t) &= D_a \frac{\partial^2 x}{\partial z^2}(0, t) \\ &= D_a \frac{\partial}{\partial z} \frac{\partial x}{\partial z}(0, t) \\ &\simeq D_a \frac{\frac{\partial x}{\partial z}(0, t) - \frac{\partial x}{\partial z}(-\Delta z, t)}{\Delta z} \\ &\simeq \frac{D_a}{\Delta z} \frac{x(\Delta z, t) - x(0, t)}{\Delta z} \\ &= \frac{D_a}{\Delta z^2} (x(\Delta z, t) - x(0, t)) \end{aligned}$$

as  $\frac{\partial x}{\partial z}(-\Delta z, t) = 0$ . Using central differences at  $z_i$ ,  $i = 2, \dots, n-1$ , leads to

$$\begin{aligned} \dot{x}(z_i, t) &= D_a \frac{\partial^2 x}{\partial z^2}(z_i, t) \\ &\simeq D_a \frac{x(z_i + \Delta z, t) - 2x(z_i, t) + x(z_i - \Delta z, t)}{\Delta z^2} \\ &= \frac{D_a}{\Delta z^2} x(z_i - \Delta z, t) - 2 \frac{D_a}{\Delta z^2} x(z_i, t) + \frac{D_a}{\Delta z^2} x(z_i + \Delta z, t). \end{aligned}$$

Finally, at  $z_n$  we have

$$\begin{aligned}
 \dot{x}(z_n, t) &= D_a \frac{\partial^2 x}{\partial z^2}(L, t) \\
 &= D_a \frac{\partial}{\partial z} \frac{\partial x}{\partial z}(L, t) \\
 &\simeq D_a \frac{\frac{\partial x}{\partial z}(L + \Delta z, t) - \frac{\partial x}{\partial z}(L, t)}{\Delta z} \\
 &\simeq \frac{D_a}{\Delta z} \frac{-x(L, t) + x(L - \Delta z, t)}{\Delta z} \\
 &= \frac{D_a}{\Delta z^2} (x(L - \Delta z, t) - x(L, t))
 \end{aligned}$$

as  $\frac{\partial x}{\partial z}(L + \Delta z, t) = 0$ . Subsection 2.3.3 shows the convergence of  $\Delta z^{-1}$ , considered as a piecewise constant function, to the Dirac delta function. This leads to the finite-dimensional system described by the equation

$$\dot{x}^{(n)} = A^{(n)} x^{(n)} + b^{(n)} u \quad (2.2.5)$$

where

$$A^{(n)} = \begin{pmatrix} -p_2 & p_2 & 0 & \cdots & 0 \\ p_2 & -2p_2 & p_2 & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & p_2 & -2p_2 & p_2 \\ 0 & \cdots & 0 & p_2 & -p_2 \end{pmatrix} \in \mathbb{R}^{n \times n}, \quad (2.2.6)$$

$$b^{(n)} = \begin{pmatrix} p_1 & 0 & \cdots & 0 \end{pmatrix}^T \in \mathbb{R}^n, \quad (2.2.7)$$

$$x^{(n)} = \begin{pmatrix} x(z_1) & \cdots & x(z_n) \end{pmatrix}^T \in \mathbb{R}^n$$

and

$$p_1 = \frac{1}{\Delta z} \quad p_2 = \frac{D_a}{\Delta z^2}.$$

**Remark 8** Initially, another discretization method was considered. The idea is to approximate the state by a finite series

$$x^*(z, t) := \sum_{i=1}^n x(z_i, t) B_{n,i}(z)$$

where the  $z_i$  are the interpolation points and the  $B_{n,i}(\cdot)$  are the bases such that the  $i^{\text{th}}$  basis, corresponding to the  $i^{\text{th}}$  interpolation point, is equal to one at  $z_i$  and is equal to zero at the other interpolation points, i.e.

$$\begin{cases} B_{n,i}(z_i) &= 1 \\ B_{n,j}(z_i) &= 0 \quad i \neq j. \end{cases}$$

It is then possible to provide conditions on the bases (and their space derivatives) to ensure that the discretized operators  $A^{(n)}$  and  $b^{(n)}$  are Metzler and nonnegative respectively, leading to the positivity of the system. A possibility is to consider the  $B_{n,j}$  as polynomials, so that

$$B_{n,j} = \sum_{k=0}^{d_{n,j}} b_{n,j,k} z^k,$$

and the conditions become easier. Increasing the order of the polynomials then allows to ensure the existence of a solution and a better efficiency of the discretization. However, it appears the finite difference method is leading to a simpler, more manipulable discretized system. As the performance of this method has been proved for a long time, it is chosen over other methods to compute finite-dimensional approximation models and to perform numerical simulations.

Some other finite difference schemes have been considered, though this one has been chosen as it leads to a discretized model that has the desired properties corresponding to the nominal PDE system. Clearly, the finite-dimensional system (2.2.5) is positive as  $A^{(n)}$  is Metzler and  $b^{(n)}$  is nonnegative. Moreover, the infinite-dimensional system (2.2.1)-(2.2.2) is not exponentially stable (see e.g. (Curtain and Zwart 1995), (Abouzaid et al. 2010)) as zero is in the spectrum of its generator

$$A = D_a \frac{d^2}{dz^2}$$

with domain

$$\mathcal{D}(A) = \left\{ x \in H = L^2(0, L) \mid x, \frac{dx}{dz} \in H \text{ are a.c., } \frac{d^2x}{dz^2} \in H, \frac{dx}{dz}(0) = 0, \frac{dx}{dz}(L) = 0 \right\}$$

where the acronym "a.c." means "absolutely continuous", that is, for every positive number  $\varepsilon$ , there is a positive number  $\delta$  such that whenever a finite sequence of pairwise disjoint subintervals  $[a_k, b_k]$  of  $[0, L]$  with  $a_k, b_k \in [0, L]$  satisfies

$$\sum_{k=1}^{\infty} (b_k - a_k) < \delta,$$

then

$$\sum_{k=1}^{\infty} |x(b_k) - x(a_k)| < \varepsilon \quad \text{and} \quad \sum_{k=1}^{\infty} \left| \frac{dx}{dz}(b_k) - \frac{dx}{dz}(a_k) \right| < \varepsilon$$

(see e.g. (Bruckner, Bruckner and Thomson 1997, Definition 5.25)). Equivalently, a function  $F$  on  $[a, b]$  is absolutely continuous if and only if there exists a function  $f \in L^1(a, b)$  such that  $F(x) = F(a) + \int_a^x f(t)dt$  for all  $x \in [a, b]$ , i.e. an absolutely continuous function is a function for which the fundamental theorem of calculus is valid. Discretizing the system will perturb the spectrum though one easily sees that the finite-dimensional system is not exponentially stable as zero is still in the spectrum of  $A^{(n)}$ . Indeed, we can see that  $A^{(n)}$  has row entries that sum to zero. This implies the last column of  $A^{(n)}$  is a linear combination of the other ones and then  $A^{(n)}$  is not full rank, which also means the determinant is null, and then zero is eigenvalue of  $A^{(n)}$ . Also note that all eigenvalues are real,  $A^{(n)}$  being symmetric. It is illustrated by Figure 2.1, that shows the discretized system eigenvalues for  $n = 10$ .

Figures 2.2 and 2.3 show the behavior of the discretized system spectrum, for  $n = 10$  and  $n = 100$  respectively, compared to the nominal system eigenvalues. Note that in Figure 2.3, we choose to show the first ten eigenvalues only to increase clarity and to allow a better comparison with Figure 2.2. It clearly appears that increasing the number of discretization points naturally leads to a more consistent spectrum, regarding that of the nominal system.

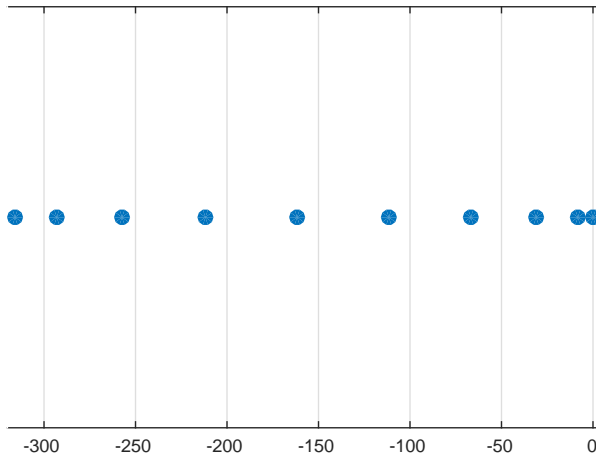


Figure 2.1: Discretized diffusion system eigenvalues ( $n = 10$ )

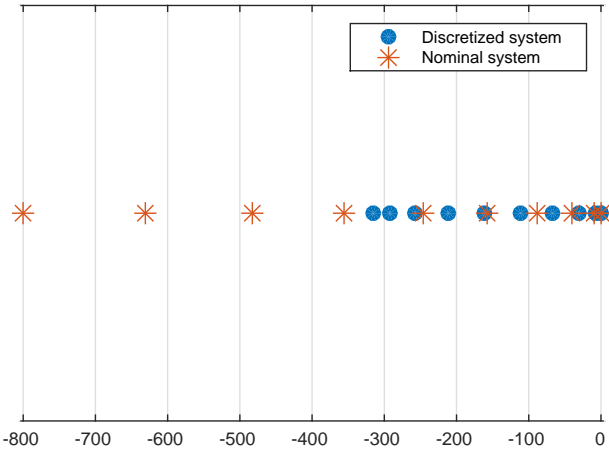


Figure 2.2: Discretized vs. nominal diffusion system eigenvalues ( $n = 10$ )

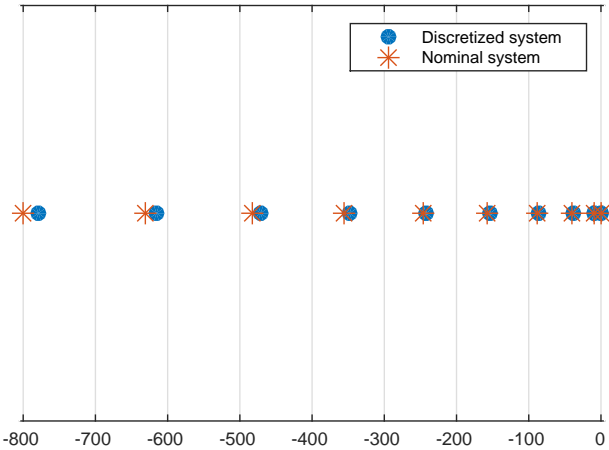


Figure 2.3: Discretized vs. nominal diffusion system eigenvalues ( $n = 100$ )

### 2.2.2 Feedback design

Our goal is thus to provide a parameterization of all positively stabilizing feedbacks for this discretized model. It appears that the system falls in the particular class we described in Theorem 1.3.2. This leads to the following result.

**Theorem 2.2.1** *A feedback matrix  $k = (k_1 \ \dots \ k_n)$  is positively stabilizing for the discretized pure diffusion system (2.2.5) if and only if it is such that*

$$k_1 = \frac{D_a v_1 - D_a v_2 - k_2 v_2 \Delta z - \dots - k_n v_n \Delta z - \Delta z^2 \omega}{v_1 \Delta z},$$

$$k_2 \geq \frac{-D_a}{\Delta z}$$

and

$$k_i \geq 0 \quad i = 3, \dots, n$$

with  $\omega > 0$  (free parameter) and such that  $v$  is a strictly positive solution of the strict inequalities set

$$\begin{aligned} -v_1 + 2v_2 - v_3 &> 0 \\ &\vdots \\ -v_{n-2} + 2v_{n-1} - v_n &> 0 \\ -v_{n-1} + v_n &> 0. \end{aligned} \tag{2.2.8}$$

In addition, a vector  $v \gg 0$  is solution of (2.2.8) if and only if it is given by one of the three following equivalent parameterizations:

1.  $\alpha$ -parameterization:  $v_n = \alpha_n$  and

$$v_{n-i} = \alpha_n - \sum_{j=1}^i (i-j+1) \alpha_{n-j} \quad i = 1, \dots, n-1$$

with  $\alpha_i > 0$  for  $i = 1, \dots, n$ , and such that

$$\alpha_n > \sum_{j=1}^{n-1} j \alpha_j$$

2.  $\beta$ -parameterization:  $v_n = \beta_n$  and

$$v_{n-i} = \beta_n - \sum_{j=1}^i \beta_{n-j} \quad i = 1, \dots, n-1$$

with  $\beta_i > 0$  for  $i = 1, \dots, n$ , and such that

$$\beta_{i+1} < \beta_i \quad i = 1, \dots, n-2$$

and

$$\beta_n > \sum_{j=1}^{n-1} \beta_j$$

3.  $\gamma$ -parameterization: with  $\gamma_i > 0$ ,  $i = 1, \dots, n$ ,

$$v = \sum_{i=1}^n \gamma_i s_i$$

and the column vectors  $s_1, \dots, s_n$  form the matrix

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \vdots & 1 & 1 & \dots & 1 \\ \vdots & \vdots & 2 & \dots & 2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 2 & \dots & n-1 \end{pmatrix}.$$

**Proof.** The first part of the result is a direct consequence of Theorem 1.3.2 and is therefore trivially obtained. Let us show the validity and the equivalence of the three parameterizations.

1. Parameterization by gap variables. Let us rewrite the strict inequalities set (2.2.8) as

$$\begin{aligned} -v_1 + 2v_2 - v_3 &= \alpha_1 \\ &\vdots \\ -v_{n-2} + 2v_{n-1} - v_n &= \alpha_{n-2} \\ -v_{n-1} + v_n &= \alpha_{n-1} \end{aligned}$$

with  $\alpha_j > 0$  for  $j = 1, \dots, n-1$ . Arbitrarily setting  $v_n = \alpha_n > 0$ , we can express  $v_{n-1} = \alpha_n - \alpha_{n-1}$  by means of the last equation. Following the same reasoning for all the other equations, we deduce that

$$v_{n-i} = \alpha_n - \sum_{j=1}^i (i-j+1)\alpha_{n-j} \quad i = 1, \dots, n-1$$

and one can see by the definition of  $v_1$  that  $v$  is a strictly positive vector if and only if

$$\alpha_n > \sum_{j=1}^{n-1} j \alpha_j.$$

2. Parameterization by finite difference. Let us define the gap variables

$$\beta_i := v_{i+1} - v_i \quad i = 1, \dots, n-1$$

and arbitrarily set  $v_n = \beta_n > 0$ . Then we can rewrite the strict inequalities set (2.2.8) as

$$\begin{aligned} \beta_1 - \beta_2 &> 0 \\ &\vdots \\ \beta_{n-2} - \beta_{n-1} &> 0 \\ \beta_{n-1} &> 0 \end{aligned}$$

and thus  $\beta_{i+1} < \beta_i$  for  $i = 1, \dots, n-2$ , with  $\beta_i > 0$  for  $i = 1, \dots, n$ . Moreover, by the definitions of  $v_n$  and  $\beta_i$ , and by the last strict inequality, we know that  $v_{n-1} = \beta_n - \beta_{n-1}$ . By recurrence

$$v_{n-i} = \beta_n - \sum_{j=1}^i \beta_{n-j} \quad i = 1, \dots, n-1$$

and one easily sees by the definition of  $v_1$  that  $v$  is a strictly positive vector if and only if

$$\beta_n > \sum_{j=1}^{n-1} \beta_j.$$

3. Parameterization by polyhedral cone vectors. Following the steps of the proof of Lemma 1.3.1, we compute  $c\{a_2^T, \dots, a_n^T, -e_1, \dots, -e_n\}$  and obtain the generating vector

$$l = \begin{pmatrix} \lambda_2 - \lambda_{n+1} \\ -2\lambda_2 + \lambda_3 - \lambda_{n+2} \\ \lambda_2 - 2\lambda_3 + \lambda_4 - \lambda_{n+3} \\ \vdots \\ \lambda_{n-2} - 2\lambda_{n-1} + \lambda_n - \lambda_{2n-1} \\ \lambda_{n-1} - \lambda_n - \lambda_{2n} \end{pmatrix}$$

for all  $\lambda_i \geq 0$ ,  $i = 2, \dots, 2n$ . It is easy to see that  $s_i^T l \leq 0$  for all  $i = 1, \dots, n$ , and thus a set of solutions of (2.2.8) is given by  $S = \mathring{c}\{s_1, \dots, s_r\}$ . The strict positivity of the  $\gamma_i$ ,  $i = 1, \dots, n$ , and the exhaustivity of the set  $S$  are deduced from the equivalence of the parameterizations, which is shown just after by means of a circular proof.

First we show that the  $\alpha$ -parameterization implies the  $\beta$ -parameterization. Consider the set of equations

$$\begin{aligned} -v_1 + 2v_2 - v_3 &= \alpha_1 \\ &\vdots \\ -v_{n-2} + 2v_{n-1} - v_n &= \alpha_{n-2} \\ -v_{n-1} + v_n &= \alpha_{n-1}. \end{aligned}$$

with  $\alpha_j > 0$ ,  $j = 1, \dots, n-1$ . By setting  $\beta_i := v_{i+1} - v_i$  for  $i = 1, \dots, n-1$ , we have that

$$\begin{aligned} \beta_1 - \beta_2 &= \alpha_1 \\ &\vdots \\ \beta_{n-2} - \beta_{n-1} &= \alpha_{n-2} \\ \beta_{n-1} &= \alpha_{n-1} \end{aligned}$$



which implies that  $\beta_{i+1} < \beta_i$  for  $i = 1, \dots, n-2$ , and  $\beta_i > 0$  for  $i = 1, \dots, n-1$ . Moreover,

$$v_n = \beta_n = \alpha_n > \sum_{j=1}^{n-1} j \alpha_j = \sum_{j=1}^{n-1} \beta_j$$

and

$$\begin{aligned} v_{n-i} &= \alpha_n - \sum_{j=1}^i (i-j+1) \alpha_{n-j} \\ &= \beta_n - \sum_{j=1}^i \beta_{n-j} \end{aligned}$$

for  $i = 1, \dots, n-1$ .

Now, let us show that the  $\beta$ -parameterization implies the  $\gamma$ -parameterization. Indeed, by setting

$$\begin{aligned} \gamma_1 &:= \beta_n - \sum_{j=1}^{n-1} \beta_j, \\ \gamma_i &:= \beta_{i-1} - \beta_i \quad i = 2, \dots, n-1 \end{aligned}$$

and

$$\gamma_n := \beta_{n-1}$$

we have that  $\gamma_i > 0$  for  $i = 1, \dots, n$ . Moreover,

$$v_n = \beta_n = \gamma_1 + \sum_{j=1}^{n-1} \beta_j = \gamma_1 + \sum_{j=1}^{n-1} j \gamma_{j+1} = \sum_{j=1}^n \gamma_j s_{j_n}$$

and

$$\begin{aligned} v_{n-i} &= \beta_n - \sum_{j=1}^i \beta_{n-j} \\ &= \gamma_1 + \sum_{j=1}^{n-1} j \gamma_{j+1} - \sum_{j=1}^i \sum_{k=1}^j \gamma_{n+1-k} \\ &= \sum_{j=1}^n \gamma_j s_{j_{n-i}} \end{aligned}$$

for  $i = 1, \dots, n-1$ .

Finally we show that the  $\gamma$ -parameterization implies the  $\alpha$ -parameterization. As

$$v = \sum_{i=1}^n \gamma_i s_i$$

we have that

$$\begin{aligned} -v_1 + 2v_2 - v_3 &= \gamma_2 \\ &\vdots \\ -v_{n-2} + 2v_{n-1} - v_n &= \gamma_{n-1} \\ -v_{n-1} + v_n &= \gamma_n \end{aligned}$$

and thus we set  $\alpha_j := \gamma_{j+1} > 0$  for  $j = 1, \dots, n-1$ . Moreover, we know that

$$\alpha_n = v_n = \gamma_1 + \sum_{j=1}^{n-1} j \gamma_{j+1} > 0$$

which implies that

$$\alpha_n - \sum_{j=1}^{n-1} j \alpha_j = \gamma_1 > 0.$$

Finally we have

$$\begin{aligned} v_{n-i} &= \gamma_1 + \sum_{j=1}^{n-1} j \gamma_{j+1} - \sum_{j=1}^i \sum_{k=1}^j \gamma_{n+1-k} \\ &= \alpha_n - \sum_{j=1}^{n-1} j \alpha_j + \sum_{j=1}^{n-1} j \alpha_j - \sum_{j=1}^i \sum_{k=1}^j \alpha_{n-k} \\ &= \alpha_n - \sum_{j=1}^i (i-j+1) \alpha_{n-j} \end{aligned}$$

and the equivalence is proved.  $\square$

**Remark 9** *Each of the parameterizations has its advantages. The  $\alpha$ -parameterization provides an intuitive approach of the problem, the  $\beta$ -parameterization provides an important piece of information concerning the structure of the vector  $v$  (a strictly increasing vector for which the gap between two successive components is strictly smaller than the previous one) and the  $\gamma$ -parameterization provides a simple and convenient expression of  $v$ .*

We can thus fully parameterize  $k_1$  and easily compute all positively stabilizing feedbacks for the discretized pure diffusion system. Note that there might be (many) more ways to parameterize the positively stabilizing feedbacks for the pure diffusion system. One of them is to consider the affine form of the Farkas lemma (see e.g. (Schrijver 1998, Corollary 7.1h), (Vivien 2002, Theorem 1)).

**Lemma 2.2.2** *Let  $\mathcal{D}$  be a non-empty polyhedron defined by  $n$  affine inequalities or faces*

$$a_k x + b_k \geq 0 \quad k = 1, \dots, n,$$

then an affine form  $\psi$  is nonnegative everywhere in  $\mathfrak{D}$  if and only if it is a nonnegative linear combination of the faces

$$\psi(x) \equiv \lambda_0 + \sum_{k=1}^n \lambda_k (a_k x + b_k) \quad \lambda_0, \dots, \lambda_n \geq 0.$$

This result can be used if we adapt our problem, and leads to the following theorem.

**Theorem 2.2.3** A feedback matrix  $k = \begin{pmatrix} k_1 & \dots & k_n \end{pmatrix}$  is positively stabilizing for the discretized pure diffusion system (2.2.5) if and only if it is such that

$$k_1 = \frac{D_a v_1 - D_a v_2 - k_2 v_2 \Delta z - \dots - k_n v_n \Delta z - \sum_{i=1}^{2n-1} \lambda_i a_i v}{v_1 \Delta z},$$

$$k_2 \geq \frac{-D_a}{\Delta z}$$

and

$$k_i \geq 0 \quad i = 3, \dots, n$$

where  $a_i$  denotes the  $i^{\text{th}}$  row of the matrix

$$a = \begin{pmatrix} -1 & 2 & -1 & 0 & \dots & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & 0 \\ \vdots & & & & \ddots & 2 & -1 \\ 0 & \dots & \dots & \dots & 0 & -1 & 1 \\ 1 & 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & \ddots & & & & & \vdots \\ \vdots & & \ddots & & & & \vdots \\ \vdots & & & \ddots & & & \vdots \\ \vdots & & & & \ddots & & \vdots \\ \vdots & & & & & \ddots & 0 \\ 0 & \dots & \dots & \dots & \dots & 0 & 1 \end{pmatrix} \in \mathbb{R}^{2n-1 \times n}$$

with  $\lambda_1, \dots, \lambda_{2n-1} \geq 0$  (free parameters) and such that  $v$  is a strictly positive solution of the strict inequalities set (2.2.8).

**Proof.** Consider the general feedback

$$k = \begin{pmatrix} k_1 & \cdots & k_n \end{pmatrix}$$

and compute the closed-loop matrix

$$A + bk = \begin{pmatrix} -p_2 + p_1 k_1 & p_2 + p_1 k_2 & p_1 k_3 & \cdots & \cdots & p_1 k_n \\ p_2 & -2p_2 & p_2 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ \vdots & & & p_2 & -2p_2 & p_2 \\ 0 & \cdots & \cdots & 0 & p_2 & -p_2 \end{pmatrix}$$

which has to be Metzler, that is

$$k_2 \geq \frac{-D_a}{\Delta z} \quad \text{and} \quad k_i \geq 0$$

for  $i = 3, \dots, n$ . Using Proposition 1.1.3 once again we know that if we can find a vector  $v \gg 0$  such that  $(A + bk)v \ll 0$ , i.e. such that

$$\begin{pmatrix} -p_2 + p_1 k_1 & p_2 + p_1 k_2 & p_1 k_3 & \cdots & \cdots & p_1 k_n \\ p_2 & -2p_2 & p_2 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ \vdots & & & p_2 & -2p_2 & p_2 \\ 0 & \cdots & \cdots & 0 & p_2 & -p_2 \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ v_n \end{pmatrix} \ll 0,$$

then  $A + bk$  is stable. This yields the same set of inequalities as before that we rewrite this time as

$$\begin{aligned} -(-p_2 + p_1 k_1)v_1 - (p_2 + p_1 k_2)v_2 - p_1 k_3 v_3 - \dots - p_1 k_n v_n + c &\geq 0 \\ -v_1 + 2v_2 - v_3 + b_1 &\geq 0 \\ &\vdots \\ -v_{n-2} + 2v_{n-1} - v_n + b_{n-2} &\geq 0 \\ -v_{n-1} + v_n + b_{n-1} &\geq 0 \\ v_1 + b_n &\geq 0 \\ &\vdots \\ v_n + b_{2n-1} &\geq 0 \end{aligned}$$

where  $c < 0$ , and  $b_i < 0$  for all  $i = 1, \dots, 2n-1$ . Recall that the set of inequalities

$$\begin{aligned}
 -v_1 + 2v_2 - v_3 &> 0 \\
 &\vdots \\
 -v_{n-2} + 2v_{n-1} - v_n &> 0 \\
 -v_{n-1} + v_n &> 0 \\
 v_1 &> 0 \\
 &\vdots \\
 v_n &> 0
 \end{aligned}$$

is feasible as Theorem 2.2.1 provides us with three different parameterizations of the solutions, and thus the set defines a non-empty polyhedron. We can then apply Lemma 2.2.2. We define

$$\begin{aligned}
 \psi(v) &:= -(-p_2 + p_1 k_1)v_1 - (p_2 + p_1 k_2)v_2 - p_1 k_3 v_3 - \dots - p_1 k_n v_n + c \\
 &\equiv \lambda_0 + \sum_{i=1}^{2n-1} \lambda_i (a_i v + b_i) \quad \lambda_0, \dots, \lambda_{2n-1} \geq 0
 \end{aligned}$$

where  $a_i$  denotes the  $i^{\text{th}}$  row of the matrix

$$a = \begin{pmatrix} -1 & 2 & -1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & 0 \\ \vdots & & & & \ddots & 2 & -1 \\ 0 & \cdots & \cdots & \cdots & 0 & -1 & 1 \\ 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & & & & & \vdots \\ \vdots & & \ddots & & & & \vdots \\ \vdots & & & \ddots & & & \vdots \\ \vdots & & & & \ddots & & \vdots \\ \vdots & & & & & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix} \in \mathbb{R}^{2n-1 \times n}$$

and

$$b = \begin{pmatrix} b_1 \\ \vdots \\ b_{2n-1} \end{pmatrix} \in \mathbb{R}^{2n-1}$$

and then

$$\begin{aligned}
 \psi(v) = & \lambda_0 \\
 & + \lambda_1(-v_1 + 2v_2 - v_3 + b_1) \\
 & \vdots \\
 & + \lambda_{n-2}(-v_{n-2} + 2v_{n-1} - v_n + b_{n-2}) \\
 & + \lambda_{n-1}(-v_{n-1} + v_n + b_{n-1}) \\
 & + \lambda_n(v_1 + b_n) \\
 & \vdots \\
 & + \lambda_{2n-1}(v_n + b_{2n-1}).
 \end{aligned}$$

Now if we compare the two expressions of  $\psi(v)$ , we can equal the constant terms to obtain

$$\begin{aligned}
 c &= \lambda_0 + \lambda_1 b_1 + \dots + \lambda_{2n-1} b_{2n-1} \\
 &= \lambda_0 + \sum_{i=1}^{2n-1} \lambda_i b_i
 \end{aligned}$$

which is clearly negative as  $\lambda_0$  can be set to zero and the sum is negative, and we can equal the variable terms to obtain

$$\begin{aligned}
 -(-p_2 + p_1 k_1)v_1 - (p_2 + p_1 k_2)v_2 - p_1 k_3 v_3 - \dots - p_1 k_n v_n = \\
 \begin{aligned}
 & \lambda_1(-v_1 + 2v_2 - v_3) \\
 & + \lambda_2(-v_2 + 2v_3 - v_4) \\
 & \vdots \\
 & + \lambda_{n-2}(-v_{n-2} + 2v_{n-1} - v_n) \\
 & + \lambda_{n-1}(-v_{n-1} + v_n) \\
 & + \lambda_n v_1 \\
 & \vdots \\
 & + \lambda_{2n-1} v_n
 \end{aligned}
 \end{aligned}$$

which makes sense as both the left and the right terms are positive. We can rewrite the last expression as

$$k_1 = \frac{D_a v_1 - D_a v_2 - k_2 v_2 \Delta z - \dots - k_n v_n \Delta z - \sum_{i=1}^{2n-1} \lambda_i a_i v}{v_1 \Delta z}.$$

□

The main difference with Theorem 2.2.1 is that instead of introducing a single parameter  $\omega$ , we use the parameters  $\lambda_i$ ,  $i = 1, \dots, 2n - 1$ , that come from Lemma 2.2.2. Both the parameterizations that are provided in Theorems 2.2.1 and 2.2.3 are equivalent and make use of the three subparameterizations ( $\alpha$ ,  $\beta$  and  $\gamma$ ) described in Theorem 2.2.1. Lemma 2.2.2 is a very useful and interesting result though, and it is essential that we highlight it as it also brings a solution to the problem.

### 2.2.3 Numerical simulations

Now we test the theoretical results using MATLAB and the discretized diffusion model we described in Subsection 2.2.1. For the following simulations we consider  $L = 1$  and  $D_a = 1$ . Let us set  $n = 11$  and let us choose the initial condition

$$x_0 = \begin{pmatrix} 1 & \dots & 1 \end{pmatrix}^T \in \mathbb{R}^{11}$$

which is the eigenvector corresponding to the Frobenius eigenvalue  $\lambda = 0$  and that excites the unstable mode, thus producing a constant profile trajectory. The open-loop state trajectory can be seen in Figure 2.4. Considering the initial condition  $x_0(z) = 2z^3 - 3z^2 + 1$  for the nominal system, which also verifies the boundary conditions and excites the (exponentially) unstable mode, leads to the open-loop state trajectory depicted in Figure 2.5.

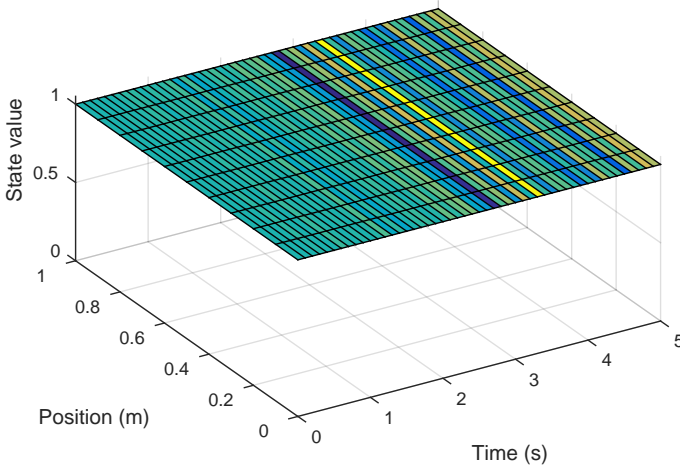


Figure 2.4: Open-loop state trajectory ( $n = 11$ ,  $x_0 = 1$ )

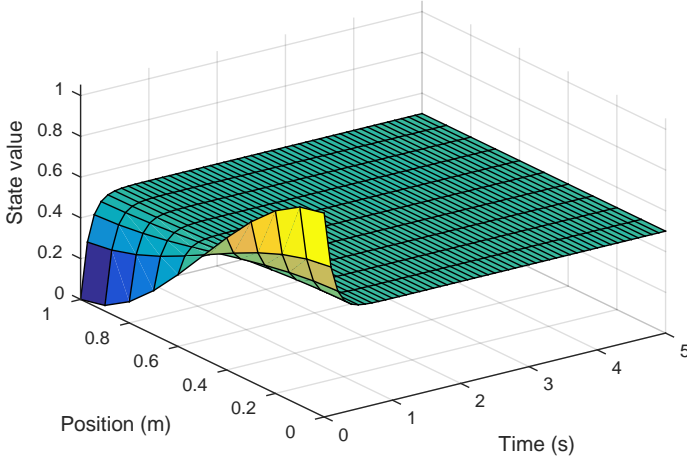


Figure 2.5: Open-loop state trajectory ( $n = 11$ ,  $x_0 = 2z^3 - 3z^2 + 1$ )

Now let us consider the feedback matrix  $k$  given by

$$k_1 = \frac{-1}{\Delta z} \kappa$$

and

$$k_i = 0 \quad i = 2, \dots, n$$

with  $\kappa \in \mathbb{R}_{0,+}$ . This particular feedback is studied in Section 2.3. Clearly, it is positively stabilizing for the discretized diffusion model as it falls in the feedback class described in Theorem 2.2.1 (see Subsection 2.3.1). Setting  $\kappa = 1$  leads to the feedback matrix

$$k = \begin{pmatrix} -10 & 0 & \dots & 0 \end{pmatrix} \in \mathbb{R}^{1 \times 11},$$

and considering the two same initial conditions as above, we obtain the input and closed-loop state trajectories as shown in Figures 2.6 and 2.7, and in Figures 2.8 and 2.9 (the MATLAB routine for the feedback design can be found in Table 2.1).



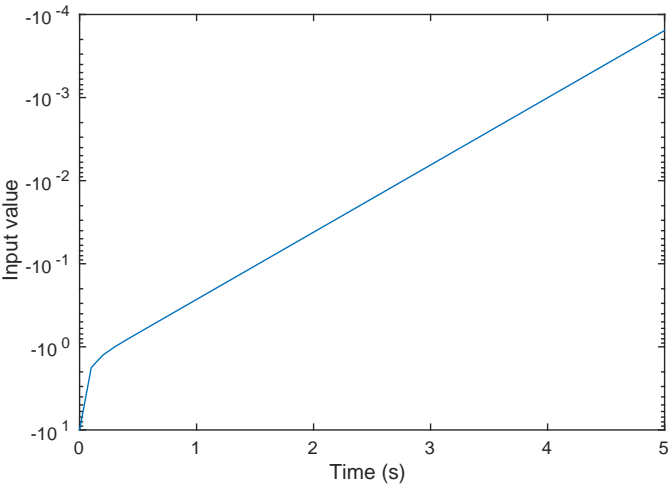


Figure 2.6: Input trajectory ( $n = 11, x_0 = 1$ )

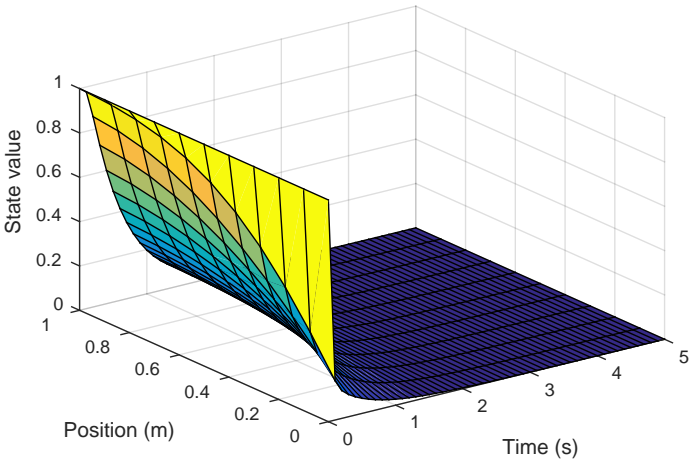


Figure 2.7: Closed-loop state trajectory ( $n = 11, x_0 = 1$ )

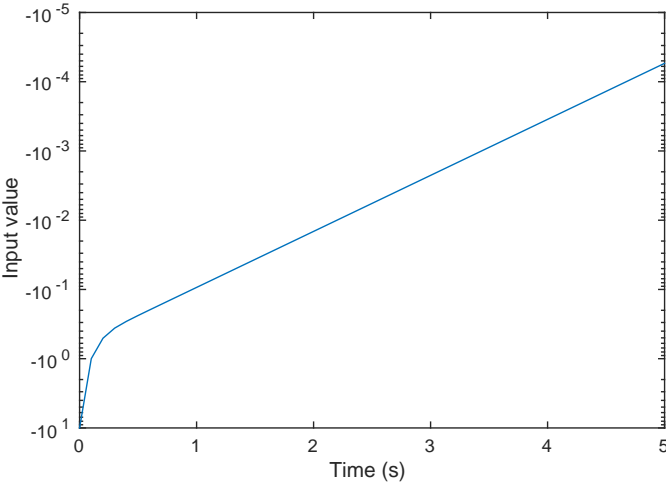


Figure 2.8: Input trajectory ( $n = 11, x_0 = 2z^3 - 3z^2 + 1$ )

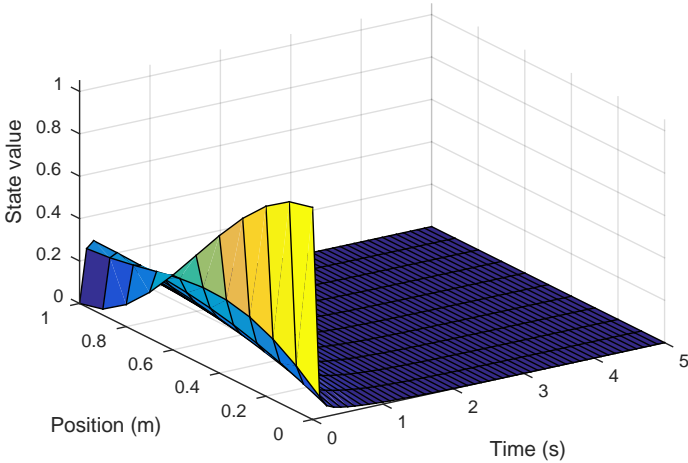


Figure 2.9: Closed-loop state trajectory ( $n = 11, x_0 = 2z^3 - 3z^2 + 1$ )

```

1 gamma = zeros(n,1);
2 for i = 1:n
3     gamma(i) = delta^(i-1);
4 end
5 s = zeros(n,n);
6 for j = 1:n
7     s(j,1) = 1;
8 end
9 for j = 2:n
10    s(j:n,j) = j-1;
11    s(j,j:n) = j-1;
12 end
13 v = zeros(n,1);
14 f = zeros(n,1);
15 for i = 1:n
16     for j = 1:n
17         f(j) = gamma(j)*s(i,j);
18     end
19     v(i) = sum(f);
20 end
21 k = zeros(1,n);
22 g = zeros(n-1,1);
23 for i = 1:n-1
24     g(i) = k(i+1)*v(i+1)*delta;
25 end
26 omega = ((Da+kappa)*v(1) - Da*v(2))/delta2;
27 k(1) = (Da*v(1)-Da*v(2)-sum(g)-delta2*omega)/(v(1)*
    delta);

```

Table 2.1: Feedback design via  $\gamma$ -parameterization

This illustrates the fact that the closed-loop system is positive and that it is stable unlike the open-loop system. So far, we have been working with  $n = 11$ . This number of discretization points might (or might not) be too weak to ensure that the simulations are reliable. Figures 2.10 and 2.11 show the comparison between the closed-loop state trajectories at  $z_n$  for different values of  $n$ , for  $x_0 = 1$  and  $x_0 = 2z^3 - 3z^2 + 1$  respectively. One can see that the difference between the curves is slight, especially when considering more than 51 discretization points.

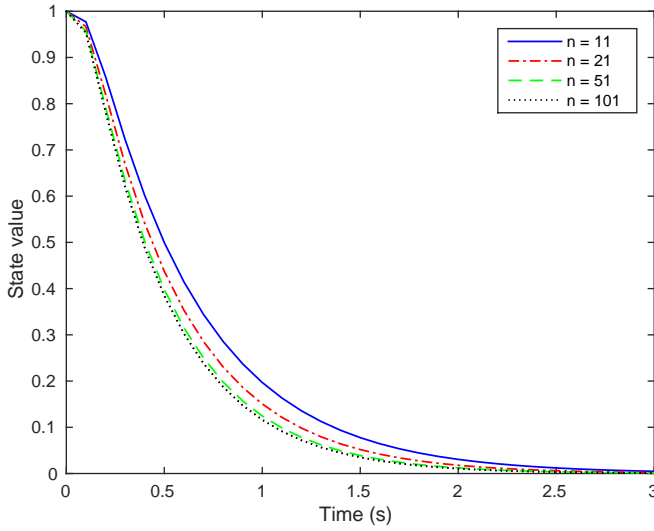
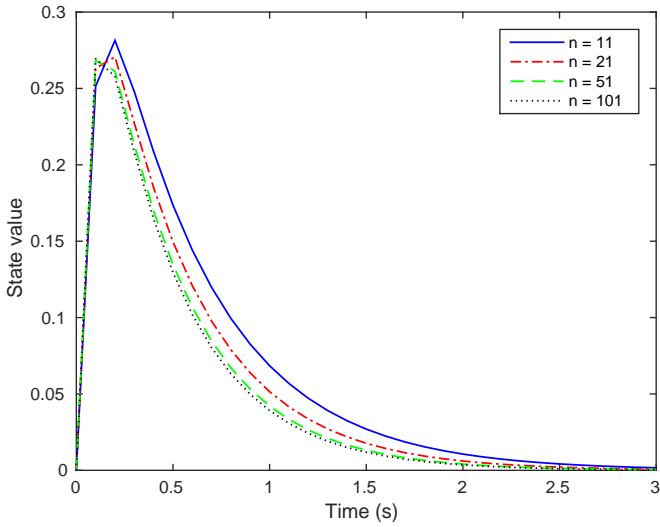
Figure 2.10: Closed-loop state trajectory at  $z_n$  ( $n = 11, 21, 51, 101, x_0 = 1$ )Figure 2.11: Closed-loop state trajectory at  $z_n$  ( $n = 11, 21, 51, 101, x_0 = 2z^3 - 3z^2 + 1$ )

Figure 2.12 provides a better understanding of the error behaviour, with respect to the number of discretization points. The blue and the red curves illustrate the absolute error defined by

$$e(i, j) = \|x^{(j)}(L) - x^{(i)}(L)\|_2$$

for  $x_0 = 1$  (see Figure 2.10) and  $x_0 = 2z^3 - 3z^2 + 1$  (see Figure 2.11), respectively. One can see that both behaviours are similar, with an expectable decreasing gain of precision.

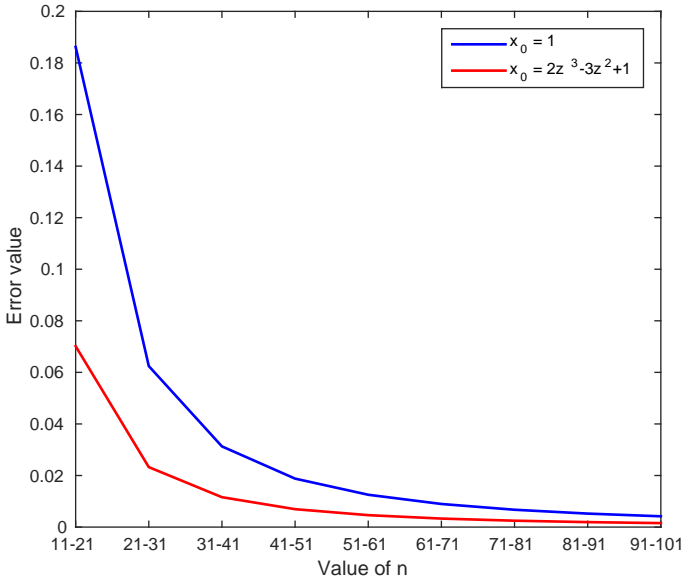
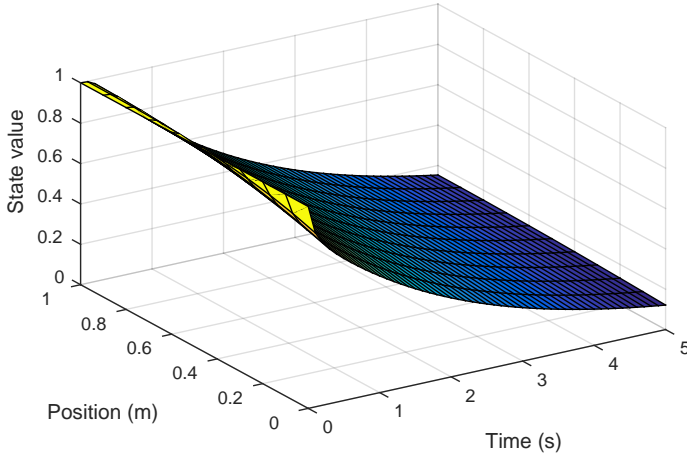
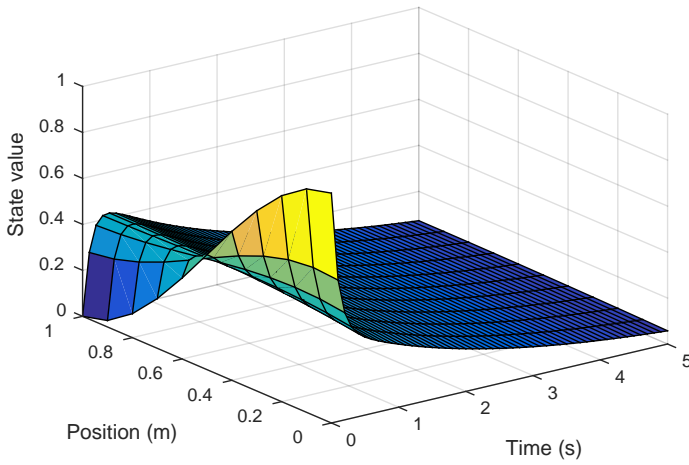


Figure 2.12: Error graph at  $z_n$  ( $n = 11, \dots, 101$ ,  $x_0 = 1$  and  $x_0 = 2z^3 - 3z^2 + 1$ )

Now let us tune the parameter  $\kappa$  and analyze the behavior of the closed-loop system. Setting  $\kappa = 0.05$  and  $\kappa = 20$ , we obtain the closed-loop state trajectories depicted in Figures 2.13 and 2.14, and in Figures 2.15 and 2.16, respectively.

Figure 2.13: Closed-loop state trajectory ( $n = 11$ ,  $\kappa = 0.05$ ,  $x_0 = 1$ )Figure 2.14: Closed-loop state trajectory ( $n = 11$ ,  $\kappa = 0.05$ ,  $x_0 = 2z^3 - 3z^2 + 1$ )

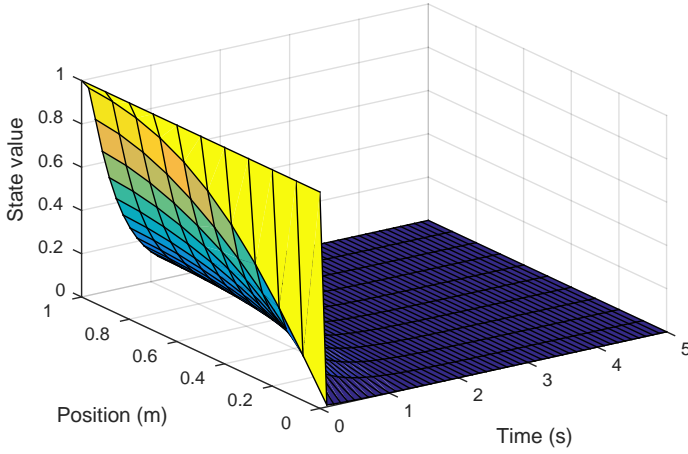


Figure 2.15: Closed-loop state trajectory ( $n = 11$ ,  $\kappa = 20$ ,  $x_0 = 1$ )

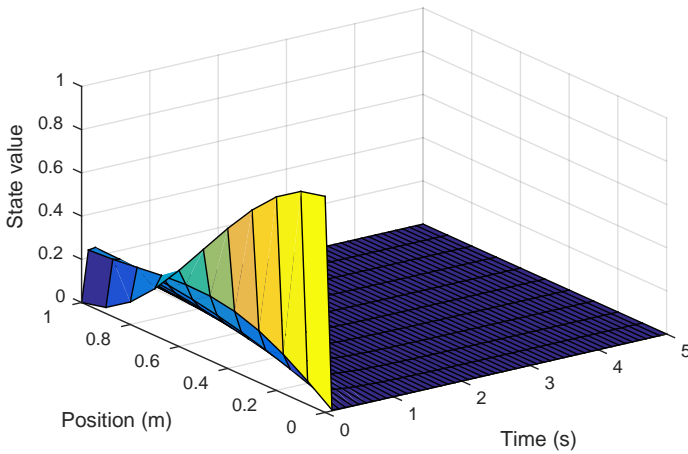


Figure 2.16: Closed-loop state trajectory ( $n = 11$ ,  $\kappa = 20$ ,  $x_0 = 2z^3 - 3z^2 + 1$ )

As expected, considering a lower value for  $\kappa$  leads to a lazy stabilization process, while a higher value for  $\kappa$  implies a higher convergence speed. Finally, Figures 2.17 and 2.18 show the step response: considering  $C = p_1 I^{(n)}$  where  $I^{(n)}$  is the identity matrix of order  $n$ , one can see that the transfer function is given by

$$\hat{G}^{(n)}(s) = p_1 w p_1$$

where  $w$  is the first column of the matrix  $(sI^{(n)} - (A^{(n)} + b^{(n)}k^{(n)}))^{-1}$ . Using the cofactor (or Laplace) expansion when computing the latter, one can show by mathematical induction that  $w$  is such that

$$w_i = \frac{\sum_{j=1}^i p_2^{n-j} C(i+j-2, 2(j-1)) s^{j-1}}{\kappa p_1^2 p_2^{n-1} + g(s)}$$

for  $i = 1, \dots, n$ , where  $C(i, j)$  is the binomial coefficient indexed by  $i$  and  $j$ , and  $g(s)$  is a polynomial function in  $s$ . The MATLAB routine which served as support for the calculations can be found in Table 2.2.

```

1  syms s p1 p2 kappa
2  n = 6;
3
4  % system matrices
5  for i = 1:n
6      vec(i) = -2*p2;
7  end
8  for i = 1:n-1
9      vec2(i) = p2;
10 end
11 A = diag(vec,0) + diag(vec2,1) + diag(vec2,-1);
12 A(1,1) = -p2;
13 A(n,n) = -p2;
14 b = zeros(n,1);
15 b(1) = p1;
16
17 % feedback and closed-loop matrix design
18 k = zeros(1,n);
19 k(1) = -kappa*p1;
20 A(1,1) = A(1,1) - kappa*p1*p1;
21
22 sol = s*eye(n)-A;
```

Table 2.2: Symbolic computation of  $(sI - (A + bk))$



The static gain is then given by

$$\hat{G}_i^{(n)}(0) = p_1 \frac{1}{\kappa p_1^2} p_1 = \frac{1}{\kappa}$$

for all  $i = 1, \dots, n$ , i.e. at every discretization points. Figures 2.19 and 2.20 show the step response at the boundaries ( $z = 0$  and  $z = 1$ ) for  $\kappa = 1$  and  $\kappa = 10$ , and one can see that the state trajectory converges faster at  $z_0$  than at  $z_n$ , which is a predictable behavior as the input only acts at  $z = 0$ . For the nominal system, one can show by means of the Laplace transform (see e.g. (Curtain and Morris 2009)) that

$$\hat{G}(s) = \frac{1}{\kappa} \left( \cosh(z\sqrt{s}) + \kappa \frac{\sinh(z\sqrt{s})}{\sqrt{s}} \right) \cosh(\sqrt{s}) - \frac{\sinh(z\sqrt{s})}{\sqrt{s}}$$

and thus

$$\hat{G}(0) = \frac{1}{\kappa} (1 + \kappa z) - z = \frac{1}{\kappa}.$$

The static gain of the discretized system, independent of the number of discretization points  $n$ , is thus equal to the static gain of the nominal system.

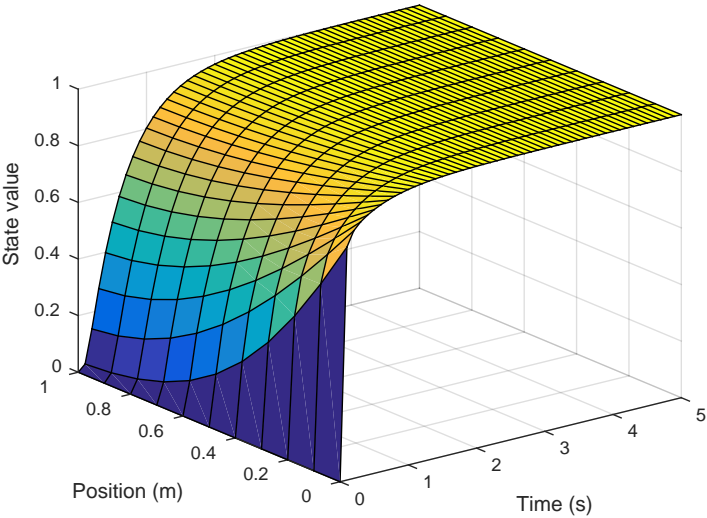


Figure 2.17: Step response ( $n = 11, \kappa = 1$ )

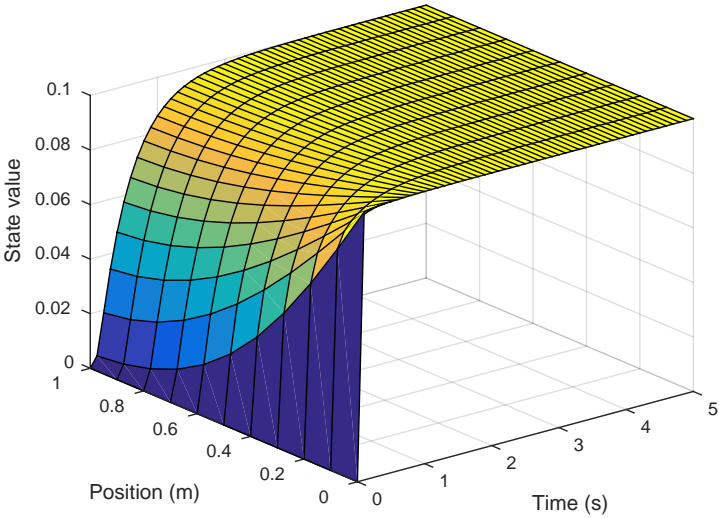
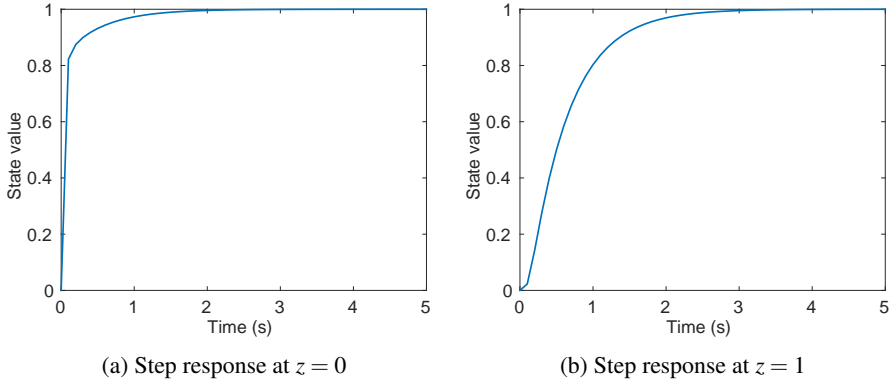
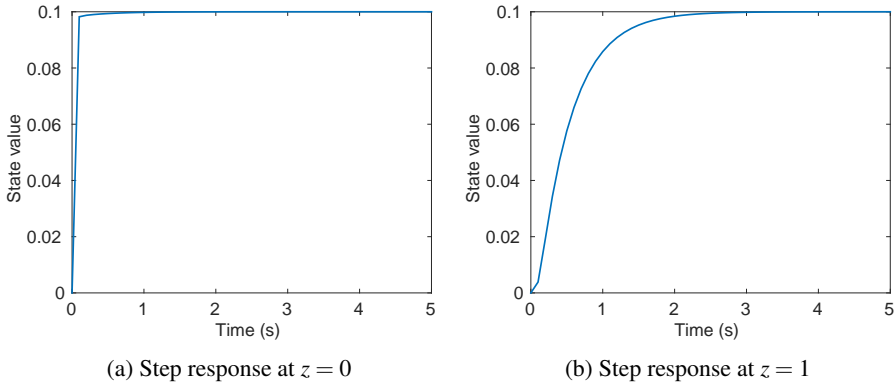


Figure 2.18: Step response ( $n = 11, \kappa = 10$ )

Figure 2.19: Step response at  $z_0$  and  $z_n$  ( $n = 11$ ,  $\kappa = 1$ )Figure 2.20: Step response at  $z_0$  and  $z_n$  ( $n = 11$ ,  $\kappa = 10$ )

## 2.3 Convergence analysis

As all positively stabilizing feedback matrices  $k^{(n)}$  for the discretized pure diffusion system were designed, we want to check the convergence of  $k^{(n)}$  to a positively stabilizing feedback operator  $k$  for the PDE system, and the convergence of the discretized closed-loop system (described by the equation  $\dot{x}^{(n)} = (A^{(n)} + b^{(n)}k^{(n)})x^{(n)}$ ) to the closed-loop PDE system (described by the equation  $\dot{x} = (A + bk)x$ ), by means of a classic study of the numerical scheme (see Subsection 2.3.2) and by means of a state space approach (see Subsection 2.3.3).

### 2.3.1 Choosing a feedback

In order to show the convergence of the closed-loop discretized system to the closed-loop PDE system, we introduce a well-chosen feedback matrix that comes from an educated guess. Most of the upcoming convergence results are independent of the input choice though, as the feedback itself is only required to show the convergence of  $k^{(n)}$  to  $k$  in the state space approach. In the following theorem, the convergence of system (2.3.2) towards system (2.3.3)-(2.3.4) (as  $n$  tends to infinity) should be understood as numerical and state space convergence, as described in Subsections 2.3.2 and 2.3.3 respectively.

**Theorem 2.3.1** *Applying the feedback  $k^{(n)}$  given by*

$$k_1^{(n)} = -\frac{1}{\Delta z} \kappa \quad \text{and} \quad k_i^{(n)} = 0 \quad (i = 2, \dots, n) \quad (2.3.1)$$

*with  $\kappa \in \mathbb{R}_{0,+}$  to the approximate system (2.2.5) leads to the convergence of the resulting closed-loop system*

$$\dot{x}^{(n)} = (A^{(n)} + b^{(n)}k^{(n)})x^{(n)}, \quad (2.3.2)$$

*as  $\Delta z$  tends to zero, to the system described by the PDE*

$$\frac{\partial x}{\partial t} = D_a \frac{\partial^2 x}{\partial z^2} \quad (2.3.3)$$

*with Neumann boundary conditions*

$$\begin{cases} \frac{\partial x}{\partial z}(0, t) &= \kappa x(0, t) \\ \frac{\partial x}{\partial z}(L, t) &= 0. \end{cases} \quad (2.3.4)$$

*Moreover, the approximate closed-loop system (2.3.2) is positive and (exponentially) stable for  $n$  sufficiently large, and the system (2.3.3)-(2.3.4) is positive and (exponentially) stable.*

**Proof.** Convergence follows from Lemmas 2.3.2 and 2.3.3. These results show the convergence of the discretized system (2.2.5) to the system described by the PDE (2.2.3), where  $u(t) = -\kappa x(0, t)$ , with homogeneous Neumann boundary conditions (2.2.4), which, by (Emirsjlow and Townley 2000, Example 2.1), is equivalent to the system (2.3.3)-(2.3.4).

Positivity of system (2.3.3)-(2.3.4), i.e. positivity of the  $C_0$ -semigroup associated with its infinitesimal generator, can be proved by standard arguments (positivity of the resolvent operator as in (Laabissi et al. 2001) or the maximum principle as in (Smith 2008)). Finally, as it is of Sturm-Liouville type, system (2.3.3)-(2.3.4) is a Riesz-spectral system (see e.g. (Delattre, Dochain and Winkin 2003)). Its spectrum is thus

real and discrete: it is easy to compute all eigenvalues and show that they are negative, implying the stability of the system.  $\square$

By Theorem 2.2.1, as  $k_i = 0$  for all  $i = 2, \dots, n$ , there holds

$$k_1^{(n)} = \frac{D_a v_1 - D_a v_2 - \Delta z^2 \omega}{v_1 \Delta z} = -\frac{1}{\Delta z} \kappa$$

where  $\omega$  should be positive. The latter condition holds if and only if

$$\frac{v_2}{v_1} < 1 + \frac{\kappa}{D_a}. \quad (2.3.5)$$

Using the  $\gamma$ -parameterization (see Theorem 2.2.1), setting  $\gamma_i = (\Delta z)^{i-1}$ , for  $i = 1, \dots, n$ , leads to a strictly positive vector  $v$  which satisfies condition (2.3.5) for  $n$  sufficiently large, as

$$\frac{v_2}{v_1} = \frac{\gamma_1 + \gamma_2 + \dots + \gamma_n}{\gamma_1} = 1 + \sum_{i=1}^n (\Delta z)^i \xrightarrow{n \rightarrow \infty} 1$$

and the constant  $\kappa/D_a$  is positive. This is illustrated in Figures 2.21 and 2.22: for  $\kappa = 1$ , the number of discretization points  $n$  should be greater or equal to three, which is pretty standard anyway, while when  $\kappa = 0.2$  one has to consider  $n = 8$  at least.

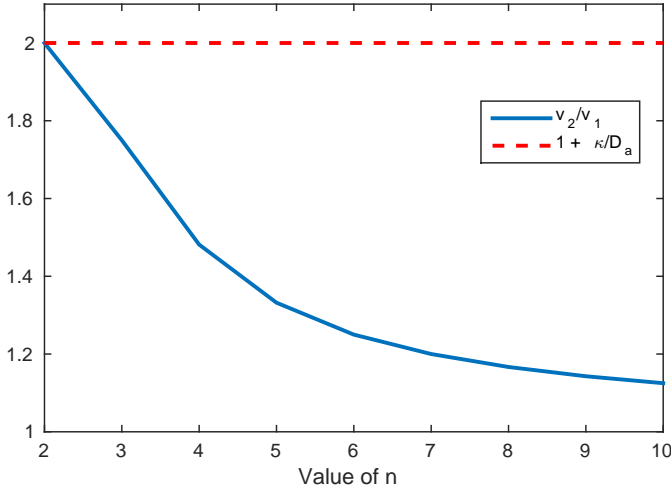
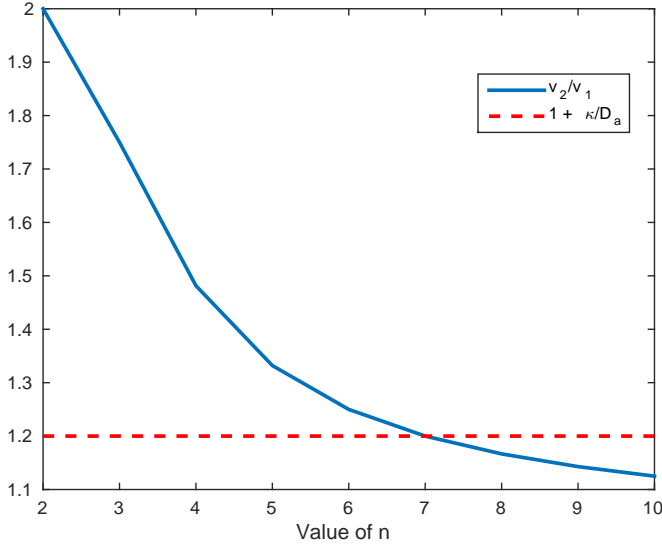


Figure 2.21: Evolution of  $v_2/v_1$  as  $n$  grows ( $\kappa = 1$ )

Figure 2.22: Evolution of  $v_2/v_1$  as  $n$  grows ( $\kappa = 0.2$ )

Equivalently, one could use the  $\alpha$ -parameterization with

$$\alpha_i = (\Delta z)^i \quad i = 1, \dots, n-1$$

and

$$\alpha_n = 1 + \Delta z + 2(\Delta z)^2 + \dots + (n-1)(\Delta z)^{n-1},$$

or the  $\beta$ -parameterization with

$$\beta_i = (\Delta z)^i + \dots + (\Delta z)^{n-1} \quad i = 1, \dots, n-1$$

and

$$\beta_n = 1 + \Delta z + 2(\Delta z)^2 + \dots + (n-1)(\Delta z)^{n-1}.$$

### 2.3.2 Consistency and stability of the scheme

Let us show the convergence of the finite difference scheme (2.2.5), independently of the choice of the input  $u$ , by proving its consistency and stability (see e.g. (Frey 2009), (Thomas 1995)). First, let us define these two concepts.

**Definition 12** A (finite difference) numerical scheme is consistent with the partial differential equation it represents if the truncation error of the scheme defined by the vector

$$\varepsilon_i^{(n)} = \mathcal{A}^{(n)}x(z_i) - f(z_i) \quad i = 1, \dots, n$$

with  $\mathcal{A}^{(n)}$  the approximation operator and  $f$  the exact one, tends uniformly towards zero with respect to  $z$ , when  $\Delta z$  tends to zero, i.e. if  $\lim_{\Delta z \rightarrow 0} \|\varepsilon^{(n)}\|_\infty = 0$ .

**Definition 13** A (finite difference) numerical scheme is said to be stable with respect to  $\|\cdot\|_\infty$  if there exists a constant  $M > 0$ , independent of the discretization step  $\Delta z$ , such that  $\|(\mathcal{A}^{(n)})^{-1}\|_\infty \leq M$ , provided that  $\mathcal{A}^{(n)}$  be invertible.

Now we can introduce the following result, which shows the convergence of the numerical scheme and which is inspired by the proof of (Frey 2009, Theorem 3.1).

**Lemma 2.3.2** If the state trajectories are sufficiently smooth, the numerical scheme developed in Section 2.2 is consistent and stable, and thus convergent.

**Proof.** Suppose that  $x$  is a  $C^4$  continuous function and extend its domain to the interval  $[-\Delta z, L + \Delta z]$  such that  $x$  is null in a small neighborhood of  $z = -\Delta z$  and  $z = L + \Delta z$ . Using Taylor expansions to the fourth order we have

$$\begin{aligned} x(z + \Delta z) &= x(z) + \Delta z x'(z) + \frac{\Delta z^2}{2} x''(z) + \frac{\Delta z^3}{6} x'''(z) + \frac{\Delta z^4}{24} x''''(\xi^+) \\ x(z - \Delta z) &= x(z) - \Delta z x'(z) + \frac{\Delta z^2}{2} x''(z) - \frac{\Delta z^3}{6} x'''(z) + \frac{\Delta z^4}{24} x''''(\xi^-) \end{aligned}$$

where  $\xi^+ \in ]z, z + \Delta z[$  and  $\xi^- \in ]z - \Delta z, z[$ . Computing the truncation error at the middle points ( $i = 2, \dots, n-1$ ) we obtain

$$\begin{aligned} \varepsilon_i^{(n)} &= D_a \frac{x(z_{i+1}) - 2x(z_i) + x(z_{i-1}))}{\Delta z^2} - D_a x''(z_i) \\ &= \frac{D_a}{\Delta z^2} \left( x(z_i) + \Delta z x'(z_i) + \frac{\Delta z^2}{2} x''(z_i) + \frac{\Delta z^3}{6} x'''(z_i) + \frac{\Delta z^4}{24} x''''(\xi_i^+) \right. \\ &\quad \left. - 2x(z_i) + x(z_i) - \Delta z x'(z_i) + \frac{\Delta z^2}{2} x''(z_i) - \frac{\Delta z^3}{6} x'''(z_i) + \frac{\Delta z^4}{24} x''''(\xi_i^-) \right) \\ &\quad - D_a x''(z_i) \\ &= D_a \frac{\Delta z^2}{12} x''''(\xi_i) \end{aligned}$$

with  $\xi_i \in ]z_{i-1}, z_{i+1}[$ . Applying a similar reasoning at the boundaries, we obtain

$$\begin{aligned} \varepsilon_1^{(n)} &= D_a \frac{x(\Delta z) - x(0)}{\Delta z^2} + \frac{1}{\Delta z} u - D_a x''(0) - \delta_0 u \\ &= D_a \frac{\Delta z^2}{12} x''''(\xi_1) \end{aligned}$$

with  $\xi_1 \in ]-\Delta z, \Delta z[$ , and

$$\begin{aligned} \varepsilon_n^{(n)} &= D_a \frac{x(L - \Delta z) - x(L)}{\Delta z^2} - D_a x''(L) \\ &= D_a \frac{\Delta z^2}{12} x''''(\xi_n) \end{aligned}$$

with  $\xi_n \in ]L - \Delta z, L + \Delta z[$ , and again (to be consistent with the discretization scheme designed in Section 2.2) we used the facts that

$$\frac{\partial x}{\partial z}(-\Delta z, t) = 0,$$

$$\frac{\partial x}{\partial z}(L + \Delta z, t) = 0$$

and  $\Delta z^{-1}$  (considered as a piecewise constant function) converges to the Dirac delta function (see Subsection 2.3.3). One easily sees that we can apply the same reasoning to

$$A_\sigma^{(n)} := A^{(n)} - \sigma I^{(n)}$$

with  $\sigma > 0$ , thus showing the consistency for the shifted system.

As  $A^{(n)}$  is not invertible we show the stability of the shifted matrix and deduce the stability of  $A^{(n)}$ . As  $A_\sigma^{(n)}$  is strictly diagonally dominant we can compute its Ahlberg-Nilson-Varah bound (see e.g. (Varah 1975, Theorem 1)) which is given by

$$\|A_\sigma^{(n)-1}\|_\infty \leq \max_{i \in \{1, \dots, n\}} \left\{ \frac{1}{|a_{ii} - \sigma| - \sum_{j=1, j \neq i}^n a_{ij}} \right\} = \frac{1}{\sigma}$$

and as it is independent of  $\Delta z$ ,  $A_\sigma^{(n)}$  is stable.

Finally we have  $A_\sigma(x - x^{(n)}) = \varepsilon_\sigma^{(n)}$ , hence

$$\|\tilde{x} - x^{(n)}\|_\infty \leq \|A_\sigma^{(n)-1}\|_\infty \|\varepsilon_\sigma^{(n)}\|_\infty$$

where  $\tilde{x} = \begin{pmatrix} x(z_1) & \dots & x(z_n) \end{pmatrix}^T \in \mathbb{R}^n$ . As the right part is bounded due to the consistency and stability properties, we have the convergence of the discretization scheme for the shifted system, and thus for the nominal one.  $\square$

### 2.3.3 State space approach

Now let us show the convergence of the system operators by a state space approach, setting the discretized operators in the appropriate spaces and using the related norms. In this section, we use the exact same notations as in (Emirsjlow and Townley 2000, Example 2.1). Let us define  $M$  as the formal system operator

$$M = \frac{d^2}{dz^2}$$



on  $L^2(0, L)$ , and let  $H = L^2(0, L)$  be the state space and  $Z = H^2(0, L)$  the domain of  $M$ . We then have  $M \in \mathcal{L}(Z, H)$ . Now let us define  $A$  as the system operator such that

$$\mathcal{D}(A) = \left\{ x \in Z \mid \frac{dx}{dz}(0, t) = \frac{dx}{dz}(L, t) = 0 \right\}$$

and  $Ax = Mx$  for  $x \in \mathcal{D}(A)$ . Then, obviously,  $A \in \mathcal{L}(Z, H)$ . A simple computation shows that the eigenvalues of  $A$  are given by

$$\lambda_k = - \left( \frac{k\pi}{L} \right)^2$$

and its eigenfunctions by

$$\begin{cases} \psi_0(z) &= \frac{1}{\sqrt{L}} \\ \psi_k(z) &= \sqrt{\frac{2}{L}} \cos\left(\frac{k\pi}{L}z\right) \end{cases} \quad k \in \mathbb{N}_0.$$

Let us define two new spaces. Firstly,  $H_1 = \mathcal{D}(A)$  equipped with the scalar product

$$\begin{aligned} \langle h, f \rangle_{H_1} &= \langle (\lambda I - A)h, (\lambda I - A)f \rangle_H \\ &= \sum_{k=0}^{\infty} (\lambda - \lambda_k)^2 a_k b_k \end{aligned}$$

and the norm

$$\|\cdot\|_{H_1} = \langle \cdot, \cdot \rangle_{H_1}^{1/2}$$

where  $h(\cdot), f(\cdot) \in \mathcal{D}(A)$ ,  $\lambda \in \mathbb{R}_+$  (arbitrary),  $a_k = \langle h, \psi_k \rangle_H$  and  $b_k = \langle f, \psi_k \rangle_H$ . Secondly,

$$H_{-1} = \left\{ x = \sum_{k=0}^{\infty} a_k \psi_k(z) \mid a_k \in \mathbb{R} \text{ and } \sum_{k=0}^{\infty} \frac{a_k^2}{(\lambda - \lambda_k)^2} < \infty \right\}$$

equipped with the scalar product

$$\begin{aligned} \langle h, f \rangle_{H_{-1}} &= \langle R(\lambda, A)h, R(\lambda, A)f \rangle_H \\ &= \sum_{k=0}^{\infty} \frac{a_k b_k}{(\lambda - \lambda_k)^2} \end{aligned}$$

and the norm

$$\|\cdot\|_{H_{-1}} = \langle \cdot, \cdot \rangle_{H_{-1}}^{1/2}.$$

Now, let us define  $U = \mathbb{R}$  as the input space and  $B \in \mathcal{L}(U, H_{-1})$  such that

$$\begin{aligned} B : \mathbb{R} &\rightarrow H_{-1} \\ u &\mapsto Bu = bu \end{aligned}$$

where the control operator  $b \in H_{-1}$  is such that  $bu = \delta_0 u$ .

We also define  $K \in \mathcal{L}(Z, \mathbb{R})$  such that

$$\begin{aligned} K : Z &\rightarrow \mathbb{R} \\ x &\mapsto Kx = \langle k, x \rangle \end{aligned}$$

where the feedback operator  $k \in H_{-1}$  is such that  $\langle k, x \rangle = \langle -\kappa \delta_0, x \rangle = -\kappa x(0)$ , with  $\kappa \in \mathbb{R}_{0,+}$ .

One should note that  $H_1 \subset H \subset H_{-1}$ , and that although the operators  $b$  and  $k$  are not bounded in  $H$ , they are bounded in  $H_{-1}$ .

Finally, for convenience, let us define

$$A_n = P_n A^{(n)} Q_n \quad (2.3.6)$$

with  $A^{(n)}$  given by (2.2.6), where

$$\begin{aligned} Q_n : Z &\rightarrow \mathbb{R}^n \\ x &\mapsto x^{(n)} = (x_1, \dots, x_n) \end{aligned}$$

and

$$\begin{aligned} P_n : \mathbb{R}^n &\rightarrow H \\ y^{(n)} &\mapsto \sum_{i=1}^{n-1} y_i^{(n)} \chi_{[(i-1)\Delta z, i\Delta z]}(\cdot) + y_n^{(n)} \chi_{\{(n-1)\Delta z\}}. \end{aligned}$$

We can now state and prove the following result.

**Lemma 2.3.3** *Let us consider the pure diffusion system (2.3.3)-(2.3.4), or equivalently*

$$\frac{\partial x}{\partial t} = D_a \frac{\partial^2 x}{\partial z^2} - \delta_0 \kappa x(0)$$

*with homogeneous Neumann boundary conditions (2.2.4), and suppose that  $x$  is a  $C^4$  continuous function. Then*

$$\begin{aligned} \lim_{n \rightarrow \infty} \|b^{(n)} - b\|_{H_{-1}} &= 0, \\ \lim_{n \rightarrow \infty} \|A_n x - Ax\|_{H_{-1}} &= 0 \quad \forall x \in Z \end{aligned}$$

and

$$\lim_{n \rightarrow \infty} \|k^{(n)} - k\|_{H_{-1}} = 0$$

*with  $b^{(n)}$ ,  $k^{(n)}$  and  $A_n$  given by (2.2.7), (2.3.1) and (2.3.6) respectively.*

**Proof.** First, we show the convergence of  $b^{(n)}$  to  $b$  (in  $H_{-1}$ ). Consider  $b^{(n)}$  as a piecewise constant function. We have

$$\|b^{(n)} - b\|_{H_{-1}}^2 = \sum_{k=0}^{\infty} \frac{a_k^2}{(\lambda - \lambda_k)^2}$$

where  $a_k = \langle b_n - b, \psi_k \rangle_H$ . Thus

$$\begin{aligned} a_0 &= \int_0^L (b^{(n)} - b)(z) \psi_0(z) dz \\ &= \int_0^L b^{(n)}(z) \psi_0(z) dz - \int_0^L b(z) \psi_0(z) dz \\ &= \int_0^{\frac{L}{n-1}} \frac{n-1}{L} \psi_0(z) dz - \psi_0(0) \\ &= \frac{1}{\sqrt{L}} - \frac{1}{\sqrt{L}} \\ &= 0 \end{aligned}$$

and, for  $k \in \mathbb{N}_0$ ,

$$\begin{aligned} a_k &= \int_0^L (b^{(n)} - b)(z) \psi_k(z) dz \\ &= \int_0^L b^{(n)}(z) \psi_k(z) dz - \int_0^L b(z) \psi_k(z) dz \\ &= \int_0^{\frac{L}{n-1}} \frac{n-1}{L} \sqrt{\frac{2}{L}} \cos\left(\frac{k\pi}{L} z\right) dz - \psi_k(0) \\ &= \sqrt{\frac{2}{L}} \left( \frac{n-1}{k\pi} \sin\left(\frac{k\pi}{n-1}\right) \right) - \sqrt{\frac{2}{L}} \end{aligned}$$

and then

$$\lim_{n \rightarrow \infty} \|b_n - b\|_{H_{-1}}^2 = \lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} \frac{\left( \sqrt{\frac{2}{L}} \left( \frac{n-1}{k\pi} \sin\left(\frac{k\pi}{n-1}\right) \right) - \sqrt{\frac{2}{L}} \right)^2}{\left( \lambda + \frac{k^2 \pi^2}{L^2} \right)^2}$$

where we can swap the limit and the series by the monotone convergence theorem. As

$$\lim_{n \rightarrow \infty} \frac{n-1}{k\pi} \sin\left(\frac{k\pi}{n-1}\right) = 1$$

the convergence of  $b^{(n)}$  to  $b$  is shown.

The convergence of  $k^{(n)}$  to  $k$  (in  $H_{-1}$ ) is straightforward as we can proceed in the same way as for  $b^{(n)}$  and  $b$  (up to a multiplicative coefficient  $-\kappa$ ), considering  $k^{(n)}$  as

a piecewise constant function.

Finally, the convergence of  $A^{(n)}$  to  $A$  is shown. Let us prove that,  $\forall x \in Z$ ,

$$\lim_{n \rightarrow \infty} \|A_n x - Ax\|_H = 0,$$

as convergence in  $H$  implies convergence in  $H_{-1}$ . First observe that

$$\begin{aligned} \|(A_n - A)x\|_H^2 &= \sum_{k=0}^{\infty} \langle (A_n - A)x, \psi_k \rangle_H^2 \\ &= \sum_{k=0}^{\infty} \int_0^L (A_n - A)x(z) \psi_k(z) dz. \end{aligned}$$

As the integral of the second term in the definition of  $P_n$  is always equal to zero, we do not consider it in the following computations. When  $k = 0$ ,

$$\begin{aligned} \int_0^L (A_n - A)x(z) \psi_0(z) dz &= \int_0^L A_n x(z) \frac{1}{\sqrt{L}} dz - \int_0^L Ax(z) \frac{1}{\sqrt{L}} dz \\ &= \frac{1}{\sqrt{L}} \sum_{i=1}^{n-1} \sum_{j=1}^n A_{ij}^{(n)} x_j^{(n)} \int_0^L \chi_{[(i-1)\Delta z, i\Delta z]}(z) dz \\ &\quad - \frac{1}{\sqrt{L}} \int_0^L x''(z) dz \\ &= \frac{1}{\sqrt{L}} \cdot 0 \cdot \Delta z - \frac{1}{\sqrt{L}} (x'(L) - x'(0)) \\ &= 0 \end{aligned}$$

and when  $k > 0$

$$\int_0^L (A_n - A)x(z) \psi_k(z) dz = \int_0^L A_n x(z) \psi_k(z) dz - \int_0^L Ax(z) \psi_k(z) dz \xrightarrow{n \rightarrow \infty} 0$$

as

$$\begin{aligned} \int_0^L A_n x(z) \psi_k(z) dz &= \sqrt{\frac{2}{L}} \sum_{i=1}^{n-1} \sum_{j=1}^n A_{ij}^{(n)} x_j^{(n)} \int_0^L \chi_{[(i-1)\Delta z, i\Delta z]}(z) \cos\left(\frac{k\pi}{L}z\right) dz \\ &= \sqrt{\frac{2}{L}} \sum_{i=1}^{n-1} \sum_{j=1}^n A_{ij}^{(n)} x_j^{(n)} \int_{(i-1)\Delta z}^{i\Delta z} \cos\left(\frac{k\pi}{L}z\right) dz \\ &= \sqrt{\frac{2}{L}} \sum_{i=1}^{n-1} \sum_{j=1}^n A_{ij}^{(n)} x_j^{(n)} \frac{L}{k\pi} \left( \sin\left(\frac{k\pi}{L}i\Delta z\right) - \sin\left(\frac{k\pi}{L}(i-1)\Delta z\right) \right) \end{aligned}$$

and

$$\begin{aligned}
 \int_0^L Ax(z)\psi_k(z)dz &= \sqrt{\frac{2}{L}} \int_0^L x''(z) \cos\left(\frac{k\pi}{L}z\right) dz \\
 &= \sqrt{\frac{2}{L}} \sum_{i=1}^{n-1} \int_{(i-1)\Delta z}^{i\Delta z} (x''(z_i) + x''(z) - x''(z_i)) \cos\left(\frac{k\pi}{L}z\right) dz \\
 &= \sqrt{\frac{2}{L}} \sum_{i=1}^{n-1} x''(z_i) \frac{L}{k\pi} \left( \sin\left(\frac{k\pi}{L}i\Delta z\right) - \sin\left(\frac{k\pi}{L}(i-1)\Delta z\right) \right) \\
 &\quad + \sqrt{\frac{2}{L}} \sum_{i=1}^{n-1} \int_{(i-1)\Delta z}^{i\Delta z} (x''(z) - x''(z_i)) \cos\left(\frac{k\pi}{L}z\right) dz
 \end{aligned}$$

where

$$\sum_{i=1}^{n-1} \int_{(i-1)\Delta z}^{i\Delta z} (x''(z) - x''(z_i)) \cos\left(\frac{k\pi}{L}z\right) dz \xrightarrow{\Delta z \rightarrow 0} 0$$

by Lebesgue's dominated convergence theorem, applied to the sequence  $(f_n)_{n \in \mathbb{N}}$  (converging pointwise to zero) of real-valued measurable functions

$$f_n(i, z) = (x''(z) - x''(z_i)) \cos\left(\frac{k\pi}{L}z\right) \chi_{[(i-1)\Delta z, i\Delta z]}(z) \chi_{\{1, \dots, n\}}(i),$$

with the dominating integrable function  $g$  given by

$$g(i, z) = \varsigma \left| \cos\left(\frac{k\pi}{L}z\right) \right| \left( \frac{L}{i-1} \right)^2$$

where  $\varsigma$  is a Lipschitz constant of the function  $x''$ , and where

$$\sum_{j=1}^n A_{ij}^{(n)} x_j^{(n)} \xrightarrow{n \rightarrow \infty} x''(z_i) \quad i = 1, \dots, n$$

thanks to the speed of convergence of the finite difference scheme (Lemma 2.3.2). Hence the convergence is proved.

The strong convergence in  $H_{-1}$  of  $A^{(n)} + b^{(n)}k^{(n)}$  to  $A + bk$ , generator of a stable and positive  $C_0$ -semigroup, is thus shown.  $\square$

## 2.4 Positive stabilization preserving boundary conditions

To conclude this chapter, let us consider how boundary conditions can be considered when one discretizes a PDE system. One could try to ensure that the state verifies the boundary conditions at all time when working with the approximate model, though they will assuredly be verified for the nominal PDE system. The goal of this section is to show how to express the boundary conditions as algebraic conditions and include them in the discretized model.

### 2.4.1 A classic approach

Consider a positive LTI system with boundary conditions, whose dynamics are described by PDEs. Discretizing the system and the boundary conditions,  $n$  being the discretized state space dimension, leads to a model of the form

$$\begin{cases} \dot{x} &= Ax + Bu \\ 0 &= Cx + Du \end{cases} \quad (2.4.1)$$

where  $x$  is the state and  $u$  is the input.

**Remark 10** *Regarding Section 2.1, one should note that in system (2.4.1) and in the following results, either  $B$  or  $D$  is null, depending on whether the input acts in the boundary conditions or in the dynamics. One should also recall that the input  $u$  is not the same depending on where it acts, so using the notation  $u$  in both cases is actually a misuse of language in order to simplify the expressions.*

In order to solve this problem, one could use the theory of singular systems (see e.g. (Burl 1985), (Dai 1989), (Lewis 1986), (Verghese, Levy and Kailath 1981)), and more specifically the theory of positive singular systems (see e.g. (Ait Rami and Napp 2012), (Shafai and Li 2016), (Virnik 2008)), considering

$$E\dot{\mathfrak{x}} = \mathcal{A}\mathfrak{x} + \mathcal{B}u$$

where

$$\mathfrak{x} = \begin{pmatrix} x \\ x \end{pmatrix}$$

and  $E$  is a singular matrix given by

$$E = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}.$$

Then one can rewrite the problem as

$$\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \dot{\mathfrak{x}} = \begin{pmatrix} A & 0 \\ 0 & C \end{pmatrix} \mathfrak{x} + \begin{pmatrix} B \\ D \end{pmatrix} u$$

which is an equivalent representation of system (2.4.1).

However, we are considering a different approach in this work. Regarding system (2.4.1) and considering a state feedback, we can rewrite the model as

$$\begin{cases} \dot{x} &= Ax + BKx = (A + BK)x \\ 0 &= Cx + DKx = (C + DK)x. \end{cases} \quad (2.4.2)$$

The idea is thus to design the feedback matrix  $K$  such that the closed-loop system is positive and stable, i.e. such that there exists a positive definite diagonal matrix  $Q$  such that  $Q(A+BK)^T + (A+BK)Q$  is a negative definite matrix (see e.g. (Boyd et al. 1994, Section 10.3)) and  $A+BK$  is Metzler, while preserving the boundary conditions at all time. To achieve the latter, one could compute the feedback matrix  $K$  such that  $C+DK=0$ , though this condition might be too strong and it would be interesting to consider a weaker condition. To summarize, we want to compute  $K$  such that

1.  $X_K := \mathbb{R}_+^n \cap \ker(C+DK)$  is  $e^{(A+BK)t}$ -invariant, and
2.  $A+BK$  is exponentially stable on  $X_K$ ,

or, equivalently, such that

$$\forall t \geq 0, \forall x_0 \in X_K, \quad x(t) := e^{(A+BK)t} x_0 \in X_K$$

and

$$\forall x_0 \in X_K, \quad e^{(A+BK)t} x_0 \xrightarrow[t \rightarrow \infty]{} 0.$$

In order to achieve this goal, let us introduce the following result.

**Theorem 2.4.1** *The set  $X_K$  is  $(A+BK)$ -invariant if and only if there exists a Metzler matrix*

$$H = \begin{pmatrix} \overset{n}{H_{11}} & \overset{p}{H_{12}} & \overset{p}{H_{13}} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{pmatrix} \begin{matrix} n \\ p \\ p \end{matrix}$$

such that

$$\begin{cases} H_{11} - H_{12}(C+DK) + H_{13}(C+DK) &= A+BK \\ H_{21} - H_{22}(C+DK) + H_{23}(C+DK) &= -(C+DK)(A+BK) \\ H_{31} - H_{32}(C+DK) + H_{33}(C+DK) &= (C+DK)(A+BK) \end{cases}$$

where  $n$  is the state space dimension and  $p$  is the number of boundary conditions.

**Remark 11** *In this result and the forthcoming ones, we use the concept of  $(A+BK)$ -invariance as it is a sufficient condition for the  $e^{(A+BK)t}$ -invariance. Indeed, we know that*

$$e^{(A+BK)t} x_0 = \sum_{l=0}^{\infty} \frac{t^l}{l!} (A+BK)^l x_0$$

and then, as  $X_K$  is closed, it is  $e^{(A+BK)t}$ -invariant if it is  $(A+BK)$ -invariant.

**Proof.** By definition,  $\ker(C+DK)$  is  $(A+BK)$ -invariant if  $(C+DK)(A+BK)x_0 = 0$  for all  $x_0$  such that  $(C+DK)x_0 = 0$ . Now, by definition of the set  $X_K$ , we know that  $x \in X_K$  if  $x \geq 0$  and  $(C+DK)x = 0$ . We can rewrite these conditions as

$$\begin{cases} -Ix &\leq 0 \\ (C+DK)x &\leq 0 \\ -(C+DK)x &\leq 0 \end{cases}$$

or equivalently

$$\begin{pmatrix} -I \\ C + DK \\ -(C + DK) \end{pmatrix} x \leq 0$$

which is of the form  $Qx \leq \rho$  (polyhedral set) with  $\rho = 0$ . Regarding (Castelan and Hennes 1993, Proposition 1),  $X_K$  is  $(A + BK)$ -invariant if and only if there exists a Metzler matrix  $H$  such that

$$\begin{cases} Q(A + BK) - HQ &= 0 \\ H\rho &\leq 0 \end{cases}$$

where the second condition is always verified as  $\rho = 0$ . We can then rewrite the first condition as

$$\begin{cases} -(A + BK) + H_{11} - H_{12}(C + DK) + H_{13}(C + DK) &= 0 \\ (C + DK)(A + BK) + H_{21} - H_{22}(C + DK) + H_{23}(C + DK) &= 0 \\ -(C + DK)(A + BK) + H_{31} - H_{32}(C + DK) + H_{33}(C + DK) &= 0 \end{cases}$$

and the conclusion holds.  $\square$

One can then deduce the two following corollaries, that both provide sufficient conditions for the  $(A + BK)$ -invariance of the set  $X_K$ .

**Corollary 2.4.2** *If  $A + BK$  is Metzler and  $(C + DK)(A + BK) = 0$ , then  $X_K$  is  $(A + BK)$ -invariant.*

**Proof.** Obvious by the definition of  $X_K$ .  $\square$

**Corollary 2.4.3** *If*

$$(A + BK)(I + 2(C + DK)^T(C + DK))^{-1} \text{ is Metzler,}$$

$$(C + DK)(A + BK) = 0$$

and

$$(A + BK)L_K(C + DK)^T = 0,$$

then  $X_K$  is  $(A + BK)$ -invariant.

**Proof.** Using the notations

$$C_K := C + DK \quad \text{and} \quad A_K := A + BK$$

we have that

$$QA_K - HQ = 0 \Leftrightarrow H = QA_K Q_l^\#$$



where

$$Q_l^\# = (Q^T Q)^{-1} Q^T$$

is the left pseudoinverse of  $Q$  which exists as  $Q$  is a full rank matrix. One can easily see that

$$Q^T Q = I + 2C_K^T C_K$$

is a positive definite matrix, and using the notations

$$L_K := (I + 2(C + DK)^T (C + DK))^{-1}$$

and

$$\begin{aligned} Q_K &:= (A + BK)(I + 2(C + DK)^T (C + DK))^{-1} \\ &= (A + BK)L_K \end{aligned}$$

we have that

$$\begin{aligned} H &= Q(A + BK)(Q^T Q)^{-1} Q^T \\ &= \begin{pmatrix} Q_K & -Q_K(C + DK)^T & Q_K(C + DK)^T \\ -(C + DK)Q_K & (C + DK)Q_K(C + DK)^T & -(C + DK)Q_K(C + DK)^T \\ (C + DK)Q_K & -(C + DK)Q_K(C + DK)^T & (C + DK)Q_K(C + DK)^T \end{pmatrix}. \end{aligned}$$

As  $H$  has to be Metzler, we deduce the following set of conditions:

$$\begin{cases} (C + DK)Q_K(C + DK)^T \text{ is a nonpositive diagonal matrix} \\ (C + DK)Q_K = 0 \\ Q_K(C + DK)^T = 0 \\ Q_K \text{ is a Metzler matrix.} \end{cases}$$

By the second condition we deduce that  $(C + DK)(A + BK) = 0$  as  $L_K$  is invertible. The first condition thus becomes  $(C + DK)Q_K(C + DK)^T = 0$ .  $\square$

Now that we have sufficient conditions for the positivity and boundary conditions to be preserved, we can deal with the positive stabilization of the system. One way to do so is to ensure that  $A + BK$  is Metzler and stable. In Corollary 2.4.2, the Metzler property is already needed so one can use this result and check the stability of the closed-loop system only, i.e. design the feedback matrix  $K$  such that for all  $\lambda \in \sigma(A + BK)$ ,  $\text{Re}(\lambda) < 0$ . Now let us introduce the concept of partial positive stabilizability (see (Achhab and Winkin 2014)).

**Definition 14** *A system (defined on an Hilbert space  $X$  with positive cone  $X_+$ ) is partially positively stabilizable if there exist a subset  $S \subset X$  and a state feedback control law  $u = Kx$  such that the resulting closed-loop system is stable and positive on  $S$ , i.e.*

$$(A + BK) \text{ is exponentially stable}$$

and

$$X_+ \cap S \text{ is } e^{(A+BK)t}\text{-invariant.}$$

It is now possible to state one last straightforward result for this subsection.

**Corollary 2.4.4** *The system is partially positively stabilizable on  $X_K$  if there exists a feedback matrix  $K$  such that*

1.  $A + BK$  is Metzler stable, and
2.  $(C + DK)(A + BK) = 0$ .

This actually means that we can (partially) positively stabilize the system in a classic way, and simply add some algebraic conditions on the design of the feedback matrix  $K$ .

### 2.4.2 Application: the pure diffusion system

In order to support the theoretical results of Subsection 2.4.1, let us consider system (2.2.3) again. Recall that the discretized model (2.2.5) is described by the equation

$$\dot{x}^{(n)} = A^{(n)}x^{(n)} + b^{(n)}u$$

where

$$A^{(n)} = \begin{pmatrix} -p_2 & p_2 & 0 & \cdots & 0 \\ p_2 & -2p_2 & p_2 & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & p_2 & -2p_2 & p_2 \\ 0 & \cdots & 0 & p_2 & -p_2 \end{pmatrix} \in \mathbb{R}^{n \times n},$$

$$b^{(n)} = \begin{pmatrix} p_1 & 0 & \cdots & 0 \end{pmatrix}^T \in \mathbb{R}^n,$$

$$x^{(n)} = \begin{pmatrix} x(z_1) & \cdots & x(z_n) \end{pmatrix}^T \in \mathbb{R}^n$$

and

$$p_1 = \frac{1}{\Delta z} \quad p_2 = \frac{D_a}{\Delta z^2}.$$

The discretized boundary conditions can then be written as  $C^{(n)}x^{(n)} = 0$  with

$$C^{(n)} = \begin{pmatrix} -p_1 & p_1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & -p_1 & p_1 \end{pmatrix} \in \mathbb{R}^{2 \times n}.$$

Now we want to find conditions on the feedback matrix  $k^{(n)}$  such that the closed-loop system preserves the discretized boundary conditions. In order to do so, we want

$\ker(C^{(n)})$  to be invariant, i.e. we want  $C^{(n)}(A^{(n)} + b^{(n)}k^{(n)})x_0 = 0$  for all  $x_0$  such that  $C^{(n)}x_0 = 0$ . Regarding Theorem 2.4.1, one has to find a Metzler matrix

$$H = \begin{pmatrix} \overset{n}{H_{11}} & \overset{2}{H_{12}} & \overset{2}{H_{13}} \\ \overset{n}{H_{21}} & \overset{2}{H_{22}} & \overset{2}{H_{23}} \\ \overset{n}{H_{31}} & \overset{2}{H_{32}} & \overset{2}{H_{33}} \end{pmatrix}$$

such that

$$\begin{cases} H_{11} - H_{12}C^{(n)} + H_{13}C^{(n)} &= (A^{(n)} + b^{(n)}k^{(n)}) \\ H_{21} - H_{22}C^{(n)} + H_{23}C^{(n)} &= -C^{(n)}(A^{(n)} + b^{(n)}k^{(n)}) \\ H_{31} - H_{32}C^{(n)} + H_{33}C^{(n)} &= C^{(n)}(A^{(n)} + b^{(n)}k^{(n)}). \end{cases} \quad (2.4.3)$$

A simple computation shows that  $C^{(n)}(A^{(n)} + b^{(n)}k^{(n)})$  is equal to

$$\begin{pmatrix} 2p_1p_2 - p_1^2k_1 & -3p_1p_2 - p_1^2k_2 & p_1p_2 - p_1^2k_3 & -p_1^2k_4 & \cdots & -p_1^2k_n \\ 0 & \cdots & 0 & -p_1p_2 & 3p_1p_2 & -2p_1p_2 \end{pmatrix}$$

and the set of equations (2.4.3) can be developed to obtain algebraic conditions on  $k^{(n)}$ . One easily sees that the first equation

$$H_{11} - H_{12}C^{(n)} + H_{13}C^{(n)} = (A^{(n)} + b^{(n)}k^{(n)})$$

is verified if  $H_{11} = (A^{(n)} + b^{(n)}k^{(n)})$ ,  $H_{12} = 0$  and  $H_{13} = 0$ . As we want the closed-loop system to be positive,  $(A^{(n)} + b^{(n)}k^{(n)})$  has to be Metzler and thus  $H_{11}$  will be Metzler too. The second equation

$$H_{21} - H_{22}C^{(n)} + H_{23}C^{(n)} = -C^{(n)}(A^{(n)} + b^{(n)}k^{(n)})$$

and the third equation

$$H_{31} - H_{32}C^{(n)} + H_{33}C^{(n)} = C^{(n)}(A^{(n)} + b^{(n)}k^{(n)})$$

lead respectively to the sets of conditions below, where  $H_{kl}^{i,j}$  denotes the  $ij^{\text{th}}$  entry of the submatrix  $H_{kl}$ .

$$\left\{ \begin{array}{ll}
 H_{21}^{1,1} + p_1 H_{22}^{1,1} - p_1 H_{23}^{1,1} & = -2p_1 p_2 + p_1^2 k_1 \\
 H_{21}^{1,2} - p_1 H_{22}^{1,1} + p_1 H_{23}^{1,1} & = 3p_1 p_2 + p_1^2 k_2 \\
 H_{21}^{1,3} & = -p_1 p_2 + p_1^2 k_3 \\
 H_{21}^{1,4} & = p_1^2 k_4 \\
 & \vdots \\
 H_{21}^{1,n-2} & = p_1^2 k_{n-2} \\
 H_{21}^{1,n-1} + p_1 H_{22}^{1,2} - p_1 H_{23}^{1,2} & = p_1^2 k_{n-1} \\
 H_{21}^{1,n} - p_1 H_{22}^{1,2} + p_1 H_{23}^{1,2} & = p_1^2 k_n \\
 H_{21}^{2,1} + p_1 H_{22}^{2,1} - p_1 H_{23}^{2,1} & = 0 \\
 H_{21}^{2,2} - p_1 H_{22}^{2,1} + p_1 H_{23}^{2,1} & = 0 \\
 H_{21}^{2,3} & = 0 \\
 & \vdots \\
 H_{21}^{2,n-3} & = 0 \\
 H_{21}^{2,n-2} & = p_1 p_2 \\
 H_{21}^{2,n-1} + p_1 H_{22}^{2,2} - p_1 H_{23}^{2,2} & = -3p_1 p_2 \\
 H_{21}^{2,n} - p_1 H_{22}^{2,2} + p_1 H_{23}^{2,2} & = 2p_1 p_2
 \end{array} \right.$$

and

$$\left\{ \begin{array}{ll} H_{31}^{1,1} + p_1 H_{32}^{1,1} - p_1 H_{33}^{1,1} & = 2p_1 p_2 - p_1^2 k_1 \\ H_{31}^{1,2} - p_1 H_{32}^{1,1} + p_1 H_{33}^{1,1} & = -3p_1 p_2 - p_1^2 k_2 \\ H_{31}^{1,3} & = p_1 p_2 - p_1^2 k_3 \\ H_{31}^{1,4} & = -p_1^2 k_4 \\ & \vdots \\ H_{31}^{1,n-2} & = -p_1^2 k_{n-2} \\ H_{31}^{1,n-1} + p_1 H_{32}^{1,2} - p_1 H_{33}^{1,2} & = -p_1^2 k_{n-1} \\ H_{31}^{1,n} - p_1 H_{32}^{1,2} + p_1 H_{33}^{1,2} & = -p_1^2 k_n \\ H_{31}^{2,1} + p_1 H_{32}^{2,1} - p_1 H_{33}^{2,1} & = 0 \\ H_{31}^{2,2} - p_1 H_{32}^{2,1} + p_1 H_{33}^{2,1} & = 0 \\ H_{31}^{2,3} & = 0 \\ & \vdots \\ H_{31}^{2,n-3} & = 0 \\ H_{31}^{2,n-2} & = -p_1 p_2 \\ H_{31}^{2,n-1} + p_1 H_{32}^{2,2} - p_1 H_{33}^{2,2} & = 3p_1 p_2 \\ H_{31}^{2,n} - p_1 H_{32}^{2,2} + p_1 H_{33}^{2,2} & = -2p_1 p_2 \end{array} \right.$$

with  $H_{22}$  and  $H_{33}$  Metzler, and  $H_{21}$ ,  $H_{23}$ ,  $H_{31}$  and  $H_{32}$  nonnegative. Now one can see that one of the equations of the second set, namely  $H_{31}^{2,n-2} = -p_1 p_2$ , is impossible to solve as  $H_{31}$  has to be nonnegative and  $-p_1 p_2 < 0$ . This comes from the fact that the input only acts on one extremity of the domain and not on the other: it is then not surprising that one cannot ensure the upkeep of the corresponding boundary condition.

Some additional input is needed in order to ensure that both boundary conditions are verified at all time. In order to go further into the problem, let us ignore the second boundary condition and let us focus on the first one only. Then we have to find a Metzler matrix

$$H = \begin{pmatrix} n & 1 & 1 \\ H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{pmatrix} \begin{matrix} n \\ 1 \\ 1 \end{matrix}$$

such that the sets of conditions

$$\left\{ \begin{array}{lcl} H_{21}^1 + p_1 H_{22} - p_1 H_{23} & = & -2p_1 p_2 + p_1^2 k_1 \\ H_{21}^2 - p_1 H_{22} + p_1 H_{23} & = & 3p_1 p_2 + p_1^2 k_2 \\ H_{21}^3 & = & -p_1 p_2 + p_1^2 k_3 \\ H_{21}^4 & = & p_1^2 k_4 \\ & \vdots & \\ H_{21}^n & = & p_1^2 k_n \end{array} \right. \quad (2.4.4)$$

and

$$\left\{ \begin{array}{lcl} H_{31}^1 + p_1 H_{32} - p_1 H_{33} & = & 2p_1 p_2 - p_1^2 k_1 \\ H_{31}^2 - p_1 H_{32} + p_1 H_{33} & = & -3p_1 p_2 - p_1^2 k_2 \\ H_{31}^3 & = & p_1 p_2 - p_1^2 k_3 \\ H_{31}^4 & = & -p_1^2 k_4 \\ & \vdots & \\ H_{31}^n & = & -p_1^2 k_n \end{array} \right. \quad (2.4.5)$$

are verified, with  $H_{21}$ ,  $H_{23}$ ,  $H_{31}$  and  $H_{32}$  nonnegative. Obviously, one has to set  $k_3 = \frac{p_2}{p_1}$  and  $k_i = 0$  for all  $i = 4, \dots, n$ . It is then possible to find solutions to these sets of equations, though it quickly appears that these solutions are not compatible with Theorem 2.2.1, i.e. one cannot expect to stabilize the system and to preserve the discretized boundary conditions at the same time.

For example, let us set  $k_1 = 0$ ,  $k_2 = \frac{-p_2}{p_1}$ ,  $k_3 = \frac{p_2}{p_1}$  and  $k_i = 0$  for all  $i = 4, \dots, n$ . This feedback matrix  $k$  is solution of the sets of equations (2.4.4) and (2.4.5). Moreover, it ensures the positivity of the system as the closed-loop matrix is Metzler (see Theorem 2.2.1). However, by Theorem 2.2.1

$$k_1 = \frac{D_a v_1 - D_a v_2 - k_2 v_2 \Delta z - \dots - k_n v_n \Delta z - \Delta z^2 \omega}{v_1 \Delta z}$$

which implies in this case that

$$0 = \frac{D_a v_1 - D_a v_2 + D_a v_2 - D_a v_3 - \Delta z^2 \omega}{v_1 \Delta z}$$

and then

$$\omega = \frac{D_a v_1 - D_a v_3}{\Delta z^2}$$

which obviously cannot be positive considering the structure of  $v$  (see Remark 9). Figures 2.23 and 2.24 show the open-loop and the closed-loop trajectories respectively, with initial condition  $x_0 = 2z^3 - 3z^2 + 1$ .

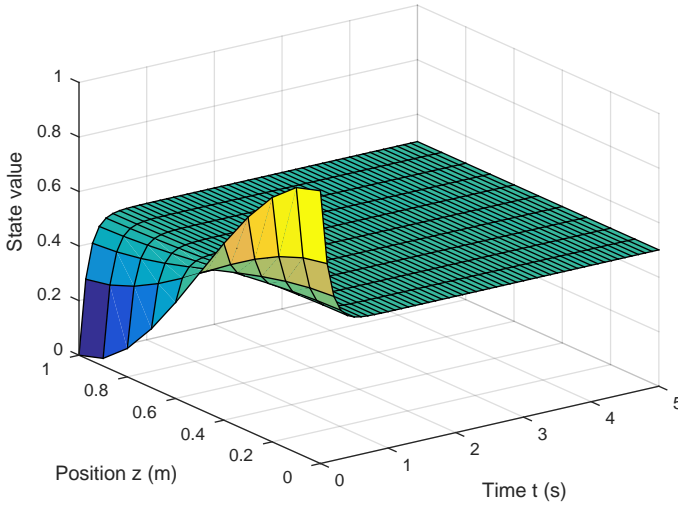


Figure 2.23: Open-loop state trajectory ( $n = 11$ ,  $x_0 = 2z^3 - 3z^2 + 1$ )

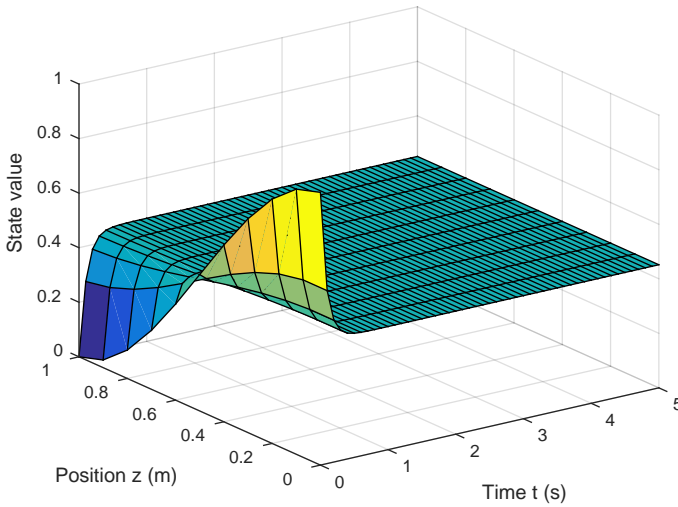


Figure 2.24: Closed-loop state trajectory ( $n = 11$ ,  $x_0 = 2z^3 - 3z^2 + 1$ )

The closed-loop system is clearly not (exponentially) stable. However, if we check the following open-loop state values

$$\begin{array}{ccc}
 x(z_1) = \begin{pmatrix} 1.0000 \\ 0.7271 \\ 0.6010 \\ 0.5449 \\ 0.5200 \\ 0.5089 \\ 0.5040 \\ 0.5018 \\ 0.5008 \\ 0.5003 \\ 0.5002 \\ 0.5001 \\ 0.5000 \\ \vdots \\ 0.5000 \end{pmatrix} & x(z_2) = \begin{pmatrix} 0.9720 \\ 0.7087 \\ 0.5928 \\ 0.5413 \\ 0.5184 \\ 0.5082 \\ 0.5036 \\ 0.5016 \\ 0.5007 \\ 0.5003 \\ 0.5001 \\ 0.5001 \\ 0.5001 \\ 0.5000 \\ \vdots \\ 0.5000 \end{pmatrix} & x(z_3) = \begin{pmatrix} 0.8960 \\ 0.6734 \\ 0.5771 \\ 0.5343 \\ 0.5153 \\ 0.5068 \\ 0.5030 \\ 0.5013 \\ 0.5006 \\ 0.5003 \\ 0.5001 \\ 0.5001 \\ 0.5001 \\ 0.5000 \\ \vdots \\ 0.5000 \end{pmatrix}
 \end{array}$$

we can see that the boundary condition is not verified, while if we check the corresponding closed-loop state values

$$\begin{array}{ccc}
 x(z_1) = \begin{pmatrix} 1.0000 \\ 0.6338 \\ 0.5197 \\ 0.4771 \\ 0.4610 \\ 0.4550 \\ 0.4528 \\ 0.4519 \\ 0.4516 \\ 0.4515 \\ 0.4514 \\ \vdots \\ 0.4514 \end{pmatrix} & x(z_2) = \begin{pmatrix} 0.9720 \\ 0.6338 \\ 0.5197 \\ 0.4771 \\ 0.4610 \\ 0.4550 \\ 0.4528 \\ 0.4519 \\ 0.4516 \\ 0.4515 \\ 0.4514 \\ \vdots \\ 0.4514 \end{pmatrix} & x(z_3) = \begin{pmatrix} 0.8960 \\ 0.6157 \\ 0.5130 \\ 0.4745 \\ 0.4601 \\ 0.4547 \\ 0.4526 \\ 0.4519 \\ 0.4516 \\ 0.4515 \\ 0.4514 \\ \vdots \\ 0.4514 \end{pmatrix}
 \end{array}$$

it appears that the boundary condition is verified at all time (except at  $t = 0$ , which is predictable as the initial state verifies the boundary condition for the PDE system only). One should note, though, that the discretized boundary conditions might actually be too strong. Although they obviously tend to the nominal boundary conditions



when  $n$  tends to infinity, they are still pretty stiff for a fixed  $n$  as it actually corresponds to a condition of equality between  $x(z_1)$  and  $x(z_2)$ , and between  $x(z_{n-1})$  and  $x(z_n)$ . A possible way to weaken the condition would be to consider inequalities instead, as

$$\begin{pmatrix} |C_1^{(n)} x^{(n)}| \\ |C_2^{(n)} x^{(n)}| \end{pmatrix} < \begin{pmatrix} \varepsilon_1^{(n)} \\ \varepsilon_2^{(n)} \end{pmatrix}$$

where  $C_i^{(n)}$  denotes the  $i^{\text{th}}$  row of the matrix  $C^{(n)}$ , and where  $\varepsilon_1^{(n)}, \varepsilon_2^{(n)} \in \mathbb{R}_+$  tend to zero when  $n$  tends to infinity. However, the theoretical results of Subsection 2.4.1 could not be applied anymore as the discretized boundary conditions would then consist in (strict) inequalities. Some adjustments would then be needed.

## Chapter 3

# Positive nonlinear time-invariant PDE systems

### 3.1 From a positive stabilization problem to an invariant stabilization problem

Let us consider a positive nonlinear finite-dimensional system described by the equation

$$\dot{x} = f(x) \quad (3.1.1)$$

with  $x(0) = x_0$ , where  $f(x)$  is a continuously differentiable function (this property guarantees the existence and uniqueness of the solution of (3.1.1), see e.g. (Khalil 2002, Section 2.2)). Now obviously, all the previous results were designed for linear systems, so it might be useful to find a way to apply these results to nonlinear systems.

Assume that there exists an equilibrium  $x_e$  for the system (3.1.1), i.e.  $f(x_e) = 0$ . The positive stabilization problem goes as follows: we want to design a feedback matrix  $K$  such that the closed-loop system is positive and the closed-loop trajectory converges towards the equilibrium  $x_e$ . In order to do so, let us approximate the nonlinear system (3.1.1) by a linear one, by using the Jacobian matrix of  $f(x)$  (see e.g. (Khalil 2002, Sections 3.3 and 11.2)). We then obtain a linearized system described by the equation

$$\dot{x}_l = Ax_l$$

where  $x_l := x - x_e$ . This change of variable implies that the equilibrium  $x_e$  corresponds to the origin for the linearized system. Therefore, in order to ensure that the nominal (nonlinear) system is positive and that the closed-loop state trajectory converges towards the equilibrium  $x_e$ , one has to design the feedback matrix  $K$  such that the closed-loop linearized system is  $C_{\bar{x}}$ -invariant, with  $\bar{x} = -x_e$  (see Section 1.4.1), and stable.

By doing so, we fall in the class of problems described in Subsections 1.4.1, 1.4.2 and 1.4.3. One can then solve the invariant stabilization problem for the linearized system by using the previous theoretical results, and afterwards check the relation between the nonlinear system and the linearized one in order to extend the feedback matrix to the nominal nonlinear system (see e.g. (Beauthier 2011, Chapter 8)).

## 3.2 Application: a tubular biochemical reactor model

In order to illustrate Section 3.1, let us introduce a relevant application: a dynamical nonlinear model of a fixed bed tubular biochemical reactor with axial dispersion (see (Dramé et al. 2008)). Tubular biochemical reactors have been studied a lot in the control field and have motivated many research activities (see e.g. (Aksikas, Winkin and Dochain 2007), (Christofides and Daoutidis 1998), (Laabissi et al. 2001), (Laabissi, Winkin, Dochain and Achhab 2005), (Logist, Saucez, Van Impe and Vande Wouwer 2009), (Smith and Zhao 1999), (Vande Wouwer et al. 2014) and references therein, (Wang, Krstić and Bastin 1999), (Winkin et al. 2000)). The dynamics of such systems are described by nonlinear partial differential equations and the nonlinearity of the model comes from the substrate inhibition term in the model equations.

### 3.2.1 Description of the system

The model described in this section comes from (Dramé et al. 2008). It is a tubular reactor containing a living biomass that feeds on a substrate. The inlet substrate is injected into the reactor via its first extremity, and diffuses within the reactor (see Figure 3.1). The reaction is considered autocatalytic, which means that the biomass is not only a product of the reaction but also a catalyst of that reaction.

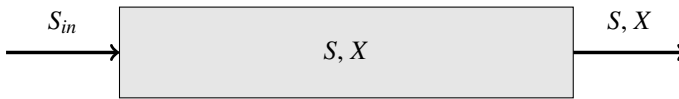


Figure 3.1: Scheme of the tubular biochemical reactor

Applying the mass balance principles to the limiting substrate concentration  $S(z, t)$  and the living biomass concentration  $X(z, t)$  leads to the positive nonlinear system described by the partial differential equations

$$\begin{cases} \frac{\partial S}{\partial t} &= D \frac{\partial^2 S}{\partial z^2} - v \frac{\partial S}{\partial z} - k\mu(S, X)X \\ \frac{\partial X}{\partial t} &= -k_d X + \mu(S, X)X \end{cases} \quad (3.2.1)$$

with boundary conditions

$$\begin{cases} D \frac{\partial S}{\partial z}(0, t) - vS(0, t) + vS_{in} &= 0 \\ \frac{\partial S}{\partial z}(L, t) &= 0 \end{cases} \quad (3.2.2)$$

for all  $t \geq 0$ . The substrate inhibition is expressed by the law

$$\mu(S, X) = \mu_0 \frac{S}{K_S X + S + \frac{1}{K_i} S^2}.$$

**Remark 12** *Note that this law describes a variant of Haldane kinetics as  $S^2$  appears in the denominator, but also  $X$ . This implies that when  $S$  increases,  $\mu$  tends to zero. In other words, feeding the biomass too much makes its mortality prevail. When dealing with Monod kinetics there is no  $S^2$  term in the denominator, implying that  $\mu$  tends to one when  $S$  tends to infinity and thus that the model does not take account of a potential overfeeding of the biomass.*

In the previous equations,  $D$ ,  $k$ ,  $k_d$ ,  $K_S$ ,  $K_i$ ,  $S_{in}$ ,  $v$  and  $\mu_0$  are all positive parameters:  $D$ ,  $v$  and  $k_d$  denote the axial dispersion coefficient, the superficial fluid velocity and the kinetic constant, respectively. The parameters  $k$  and  $K_S$  are dimensionless, and  $K_i$  has the dimension of a concentration. The growth rate  $\mu(S, X)$  has the dimension of the inverse of a time and  $S_{in}$  is the inlet limiting substrate concentration, so that  $0 \leq S \leq S_{in}$  due to the saturation condition. Finally,  $z \in [0, L]$  and  $t \geq 0$  denote the spatial and time variables respectively, and  $L$  denotes the length of the reactor.

One can observe in (3.2.1) that the two first terms in the substrate equation are simply diffusion and convection terms, respectively. The third term represents the consumption of the substrate. The first term of the biomass equation represents its mortality, while the second term represents its growth as it feeds on the substrate. The inlet substrate acts as in the input in the first boundary condition, as seen in (3.2.2). The well-posedness of the model is shown in (Dramé et al. 2008, Section III.A.).

The objective of this chapter is to stabilize the system around one of its equilibria, corresponding to an interesting biomass concentration profile, as suggested in the perspectives of (Dramé et al. 2008).

**Remark 13** *Depending on the application and the main objective, one could try not to optimize the biomass production, but to optimize the substrate inhibition in order to get rid of it. This could happen if, for example, one wants to wipe pollutant substances out. One could then study the model around a stable equilibrium, and deal with disturbance rejection.*

Let us first introduce an equivalent dimensionless infinite-dimensional system description. We define the new state variables

$$\tilde{x}_1 := \frac{S_{in} - S}{S_{in}} \quad \text{and} \quad \tilde{x}_2 := \frac{X}{S_{in}} \quad (3.2.3)$$

and the new spatial and time variables

$$\zeta := \frac{z}{L} \quad \text{and} \quad \tau := \frac{v}{L} t,$$

respectively. This leads to a model equivalent to (3.2.1)-(3.2.2) and described by the partial differential equations

$$\begin{cases} \frac{\partial \tilde{x}_1}{\partial \tau} &= \frac{1}{P_e} \frac{\partial^2 \tilde{x}_1}{\partial \zeta^2} - \frac{\partial \tilde{x}_1}{\partial \zeta} + k\tilde{\mu}(\tilde{x}_1, \tilde{x}_2)\tilde{x}_2 \\ \frac{\partial \tilde{x}_2}{\partial \tau} &= -\gamma\tilde{x}_2 + \tilde{\mu}(\tilde{x}_1, \tilde{x}_2)\tilde{x}_2 \end{cases}$$

with boundary conditions

$$\begin{cases} \frac{1}{P_e} \frac{\partial \tilde{x}_1}{\partial \zeta}(0, \tau) - \tilde{x}_1(0, \tau) &= 0 \\ \frac{\partial \tilde{x}_1}{\partial \zeta}(1, \tau) &= 0 \end{cases}$$

for all  $\tau \geq 0$ . The modified substrate inhibition law is given by

$$\tilde{\mu}(\tilde{x}_1, \tilde{x}_2) = \beta \frac{(1 - \tilde{x}_1)}{K_S \tilde{x}_2 + (1 - \tilde{x}_1) + \alpha(1 - \tilde{x}_1)^2}$$

where the constants  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $P_e$  (the Péclet number) are given by

$$\alpha := \frac{S_{in}}{K_i}, \quad \beta := \frac{L}{v}\mu_0, \quad \gamma := \frac{L}{v}k_d \quad \text{and} \quad P_e := \frac{vL}{D},$$

respectively. The equilibrium profiles are solutions of

$$\begin{cases} \frac{1}{P_e} \bar{x}_1'' - \bar{x}_1' + k\tilde{\mu}(\bar{x}_1, \bar{x}_2)\bar{x}_2 &= 0 \\ -\gamma\bar{x}_2 + \tilde{\mu}(\bar{x}_1, \bar{x}_2)\bar{x}_2 &= 0 \\ \frac{1}{P_e} \bar{x}_1'(0) - \bar{x}_1(0) = \bar{x}_1'(1) &= 0. \end{cases} \quad (3.2.4)$$

The existence of multiple equilibrium profiles is shown in (Dramé et al. 2008) by using a perturbation theory. More precisely, it is stated in (Dramé et al. 2008, Proposition 4.1) that there exist  $D^* > 0$  sufficiently large and  $v^* > 0$  such that for all  $D \geq D^*$  the system (3.2.4) has at least two non trivial solutions if the parameter  $v$  satisfies  $0 \leq v < v^*$ , and at least one non trivial solution if  $v = v^*$ . System (3.2.4) has then up to three distinct solutions, thus corresponding to equilibrium profiles for the initial system. One should note that for all the forthcoming simulations, the values of the parameters are the same as in (Dramé et al. 2008, Section IV.):

Parameter	Value
$S_{in}$	10
$K_i$	5
$\mu_0$	0.65
$k_d$	0.26
$D$	1.3
$L$	1
$k$	7
$K_S$	3
$v$	0.03

With this set of parameters, it can be shown that there exist three distinct equilibria. The MATLAB routine for the equilibria computation can be found in Table 3.1.

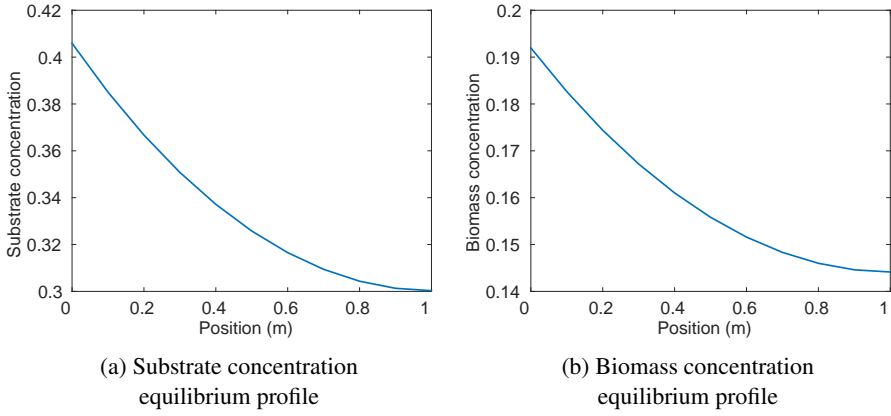
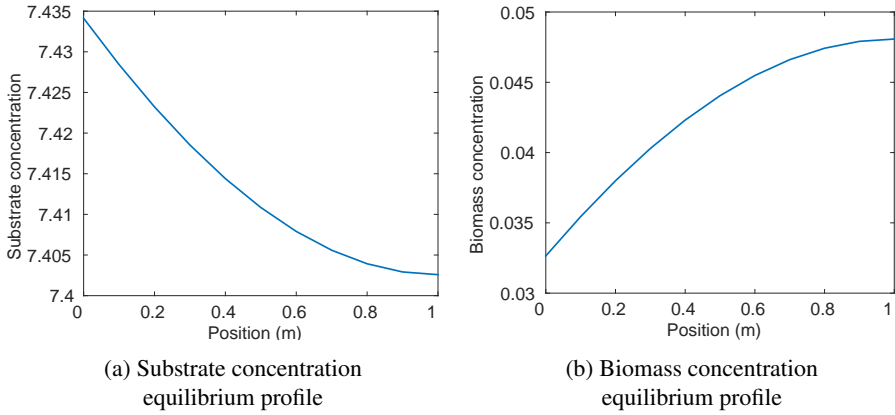
```

1 % system equations
2 function dx = biochemical(z,x)
3 dx = zeros(2,1);
4 dx(1) = x(2);
5 dx(2) = (1/D)*(nu*x(2) - (k*L*(1-x(1))*(M+alpha*kd*
        x(1)))/KS);
6
7 % boundary conditions
8 function bc = boundary(xa,xb)
9 bc = [D*xa(2) - nu*xa(1) ; xb(2)];
10
11 % solving the equations
12 solinit = bvpinit(linspace(0,1,5),[0.2 1]);
13 sol = bvp4c(@biochemical,@boundary,solinit);
14 xint = linspace(0,1,5);
15 X1 = deval(sol,xint);
16 X2 = zeros(length(X1(1,:)),1);
17 for i = 1:length(X1(1,:))
18     X2(i) = ( (1-X1(1,i))*(M+alpha*kd*X1(1,i)) ) / (
        kd*KS);
19 end

```

Table 3.1: Equilibria computation for the biochemical reactor model

The first equilibrium, stable, is the trivial solution  $(x_1, x_2) = (0, 0)$  which corresponds to the reactor washout  $(\bar{S}, \bar{X}) = (S_{in}, 0)$  with the initial state variables. The second equilibrium, denoted by  $(x_1^*, x_2^*)$ , is locally asymptotically stable and corresponds to the case where both the substrate concentration and the biomass concentration are decreasing along the reactor: the substrate concentration is too small so the death effect prevails over the growth of the biomass (see Figure 3.2). The last equilibrium, denoted by  $(\bar{x}_1, \bar{x}_2)$ , is unstable and corresponds to the case where the substrate concentration is decreasing along the reactor while the biomass concentration is increasing: the substrate concentration is large enough, leading to an "optimal" biomass concentration profile as the biomass concentration efficiently grows along the reactor (see Figure 3.3). One should note that the profiles depicted in Figures 3.2 and 3.3 are in the nominal set of variables  $(S, X)$  for a better viewing and understanding of the physical configurations.

Figure 3.2: Stable equilibrium profiles  $(x_1^*, x_2^*)$ Figure 3.3: Unstable equilibrium profiles  $(\bar{x}_1, \bar{x}_2)$ 

As we implemented a routine that numerically solves the PDEs that describe the nominal nonlinear system (the MATLAB routine can be found in Table 3.2), we can check the behavior of the (open-loop) system around the three equilibria.

```

1 % system equations
2 function [c,f,s] = PDEfunction(z,t,x,DxDz)
3 c = [1 ; 1];
4 f = [1/Pe ; 0] .* DxDz;
5 s = [-DxDz(1) + k*mu*x(2) ; -gamma*x(2) + mu*x(2)];
6
7 % boundary conditions
8 function [pl,ql,pr,qr] = PDEboundary(zl,xl,zr,xr,t)
9 pl = [-xl(1) ; 0];
10 ql = [1 ; eps];
11 pr = [0 ; 0];
12 qr = [Pe ; eps];
13
14 % initial condition
15 function x0 = PDEinit(z)
16 CI = biochemicalODE(); % equilibrium computation
17 x0 = [CI(round(z*50+1),1) ; CI(round(z*50+1),2)];
18
19 % solving system equations
20 z = linspace(0,1,51);
21 t = linspace(0,5,51);
22 m = 0;
23 sol = pdepe(m,@PDEfunction,@PDEinit,@PDEboundary,z,
24             t);
25 x1 = sol(:,:,1);
26 x2 = sol(:,:,2);

```

Table 3.2: Solving the PDE problem with *pdepe*



There is no point studying the trivial equilibrium  $(x_1, x_2) = (0, 0)$  as the biomass concentration will obviously remain null at all time. Now regarding the second equilibrium  $(x_1^*, x_2^*)$  and considering it as the initial condition of the problem, one can observe its stability in Figures 3.4 and 3.5 as the latter depict constant state profiles.

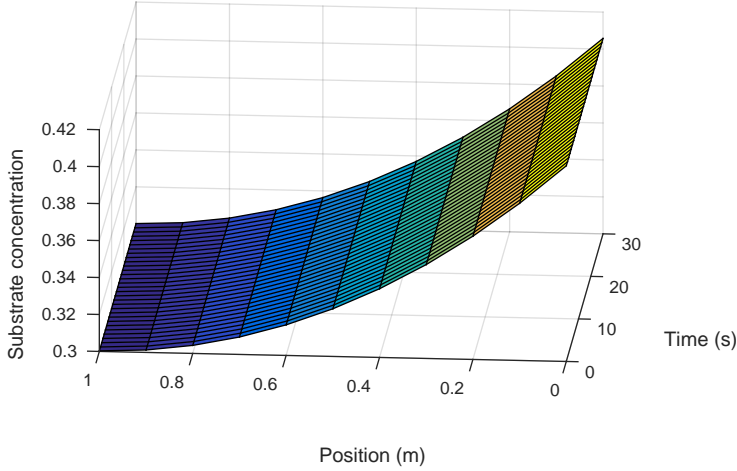


Figure 3.4: Substrate concentration trajectory ( $x_0 = (x_1^*, x_2^*)$ )

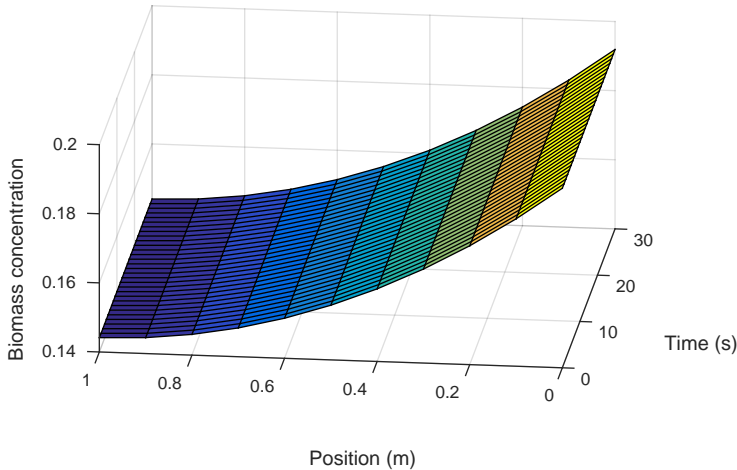


Figure 3.5: Biomass concentration trajectory ( $x_0 = (x_1^*, x_2^*)$ )

The third equilibrium  $(\bar{x}_1, \bar{x}_2)$ , the one we are interested in, is used as the initial condition in Figures 3.6 and 3.7. One can see that, though the state trajectories remain still during the first seconds, numerical errors due to the discretization and numerical integration inevitably lead to slight variations that highlight the unstability of the equilibrium.

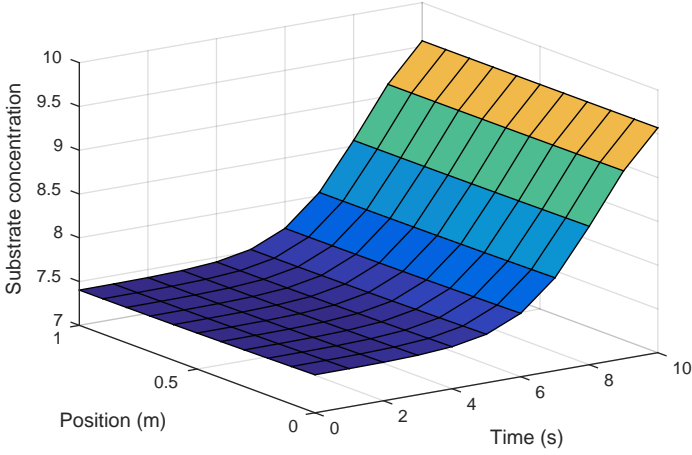


Figure 3.6: Substrate concentration trajectory ( $x_0 = (\bar{x}_1, \bar{x}_2)$ )

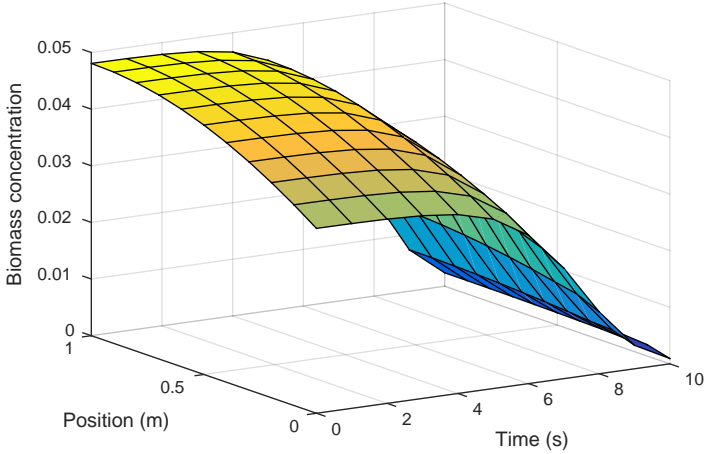


Figure 3.7: Biomass concentration trajectory ( $x_0 = (\bar{x}_1, \bar{x}_2)$ )

Figures 3.8 and 3.9 show these trajectories on a larger time interval, which also illustrate the positivity of the nominal nonlinear system, the saturation condition, and the attractivity of the trivial equilibrium.

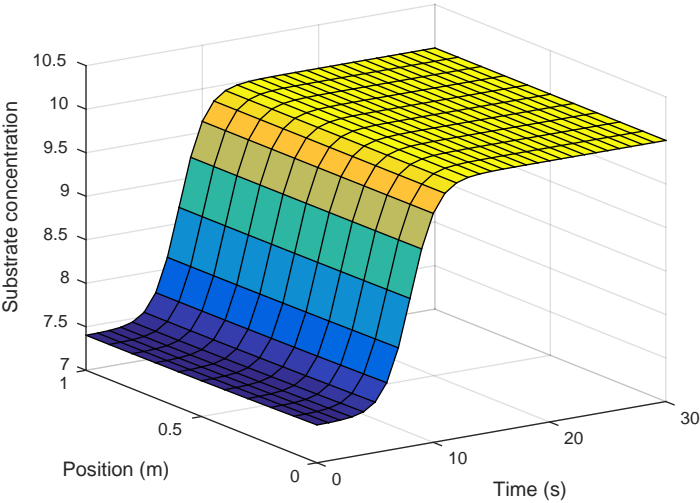


Figure 3.8: Substrate concentration trajectory ( $x_0 = (\bar{x}_1, \bar{x}_2)$ )

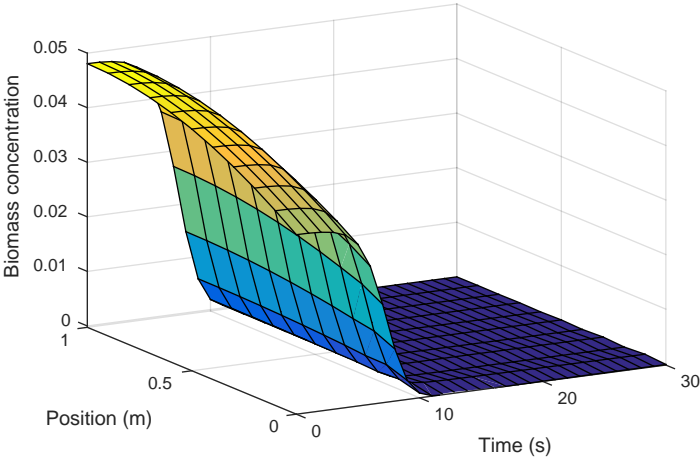


Figure 3.9: Biomass concentration trajectory ( $x_0 = (\bar{x}_1, \bar{x}_2)$ )

It is possible to manually compensate the numerical error, of course, by adding some perturbation on the initial condition. For example, in Figures 3.10 and 3.11 we choose the initial condition  $x_0 = 1.0003\bar{x}$  that obviously leads to a more static behavior.

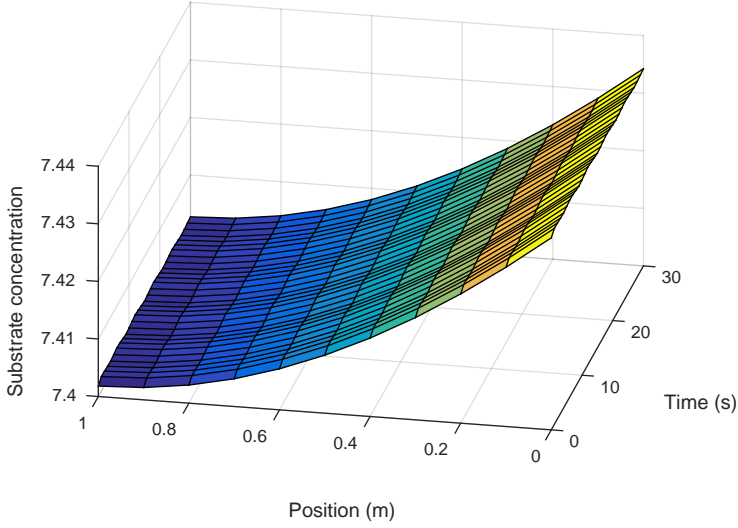


Figure 3.10: Substrate concentration trajectory ( $x_0 = 1.0003(\bar{x}_1, \bar{x}_2)$ )

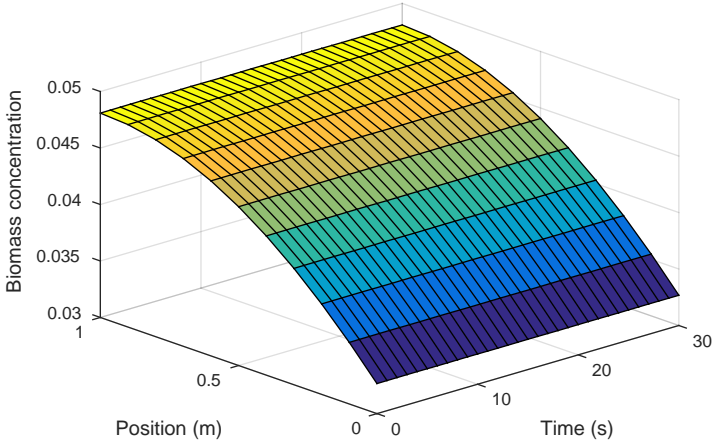


Figure 3.11: Biomass concentration trajectory ( $x_0 = 1.0003(\bar{x}_1, \bar{x}_2)$ )

Now let us go back to system (3.2.1)-(3.2.2). The objective is to stabilize the system around the unstable equilibrium to obtain the "optimal" biomass concentration profile. We will denote this desired equilibrium by  $(\bar{S}, \bar{X})$ , and regarding the change of variables (3.2.3) one can see that

$$\begin{cases} \bar{S} &= S_{in} - S_{in}\bar{x}_1 = S_{in}(1 - \bar{x}_1) \\ \bar{X} &= S_{in}\bar{x}_2. \end{cases}$$

In order to apply the theoretical results developed in this work, we linearize the system around the equilibrium. This leads to the LTI system described by the equations

$$\begin{cases} \frac{\partial \tilde{S}}{\partial t} &= D \frac{\partial^2 \tilde{S}}{\partial z^2} - v \frac{\partial \tilde{S}}{\partial z} - k \frac{c_1}{c_3} \tilde{S} - k \frac{c_2}{c_3} \tilde{X} \\ \frac{\partial \tilde{X}}{\partial t} &= \frac{c_1}{c_3} \tilde{S} + \left( \frac{c_2}{c_3} - k_d \right) \tilde{X} \end{cases}$$

with boundary conditions

$$\begin{cases} D \frac{\partial \tilde{S}}{\partial z}(0, t) - v \tilde{S}(0, t) + v \tilde{S}_{in} &= 0 \\ \frac{\partial \tilde{S}}{\partial z}(L, t) &= 0 \end{cases}$$

where

$$\tilde{S} := S - \bar{S}, \quad \tilde{X} := X - \bar{X}, \quad \tilde{S}_{in} := S_{in} - \bar{S}_{in}$$

and

$$\begin{aligned} c_1 &= \mu_0 \left( K_S \bar{X}^2 - \frac{1}{K_i} \bar{X} \bar{S}^2 \right), \\ c_2 &= \mu_0 \left( \bar{S}^2 + \frac{1}{K_i} \bar{S}^3 \right), \\ c_3 &= \left( K_S \bar{X} + \bar{S} + \frac{1}{K_i} \bar{S}^2 \right)^2. \end{aligned}$$

Discretizing this system by the finite difference method, adopting a similar scheme as in Section 2.2, considering  $n$  discretization points  $z_i, i = 1, \dots, n$ , with  $z_1 = 0, z_n = L$  and  $\Delta z = L/(n-1)$  the discretization step, leads to the finite-dimensional system described by the equation

$$\dot{x}^{(n)} = A^{(n)} x^{(n)} + b^{(n)} u \quad (3.2.5)$$

where

$$A^{(n)} = \begin{pmatrix} A_1^{(n)} & A_2^{(n)} \\ A_3^{(n)} & A_4^{(n)} \end{pmatrix} \in \mathbb{R}^{2n \times 2n},$$

$$b^{(n)} = \begin{pmatrix} v & 0 & \dots & 0 \end{pmatrix}^T \in \mathbb{R}^{2n}$$

and

$$x^{(n)} = \left( S(z_1) \quad \cdots \quad S(z_n) \quad X(z_1) \quad \cdots \quad X(z_n) \right)^T \in \mathbb{R}^{2n}$$

and  $u$  is the input. The submatrices of  $A^{(n)}$  are given by

$$A_1^{(n)} = \begin{pmatrix} \frac{-D}{\Delta z^2} - \frac{v}{\Delta z} - k \frac{c_{1,1}}{c_{3,1}} & \frac{D}{\Delta z^2} & 0 & \cdots & 0 \\ \frac{D}{\Delta z^2} + \frac{v}{\Delta z} & \frac{-2D}{\Delta z^2} - \frac{v}{\Delta z} - k \frac{c_{1,2}}{c_{3,2}} & \frac{D}{\Delta z^2} & & \vdots \\ 0 & & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \frac{D}{\Delta z^2} + \frac{v}{\Delta z} & \frac{-2D}{\Delta z^2} - \frac{v}{\Delta z} - k \frac{c_{1,n-1}}{c_{3,n-1}} & \frac{D}{\Delta z^2} \\ 0 & \cdots & 0 & \frac{D}{\Delta z^2} + \frac{v}{\Delta z} & \frac{-D}{\Delta z^2} - \frac{v}{\Delta z} - k \frac{c_{1,n}}{c_{3,n}} \end{pmatrix},$$

$$A_2^{(n)} = \begin{pmatrix} -k \frac{c_{2,1}}{c_{3,1}} & & \\ & \ddots & \\ & & -k \frac{c_{2,n}}{c_{3,n}} \end{pmatrix},$$

$$A_3^{(n)} = \begin{pmatrix} \frac{c_{1,1}}{c_{3,1}} & & \\ & \ddots & \\ & & \frac{c_{1,n}}{c_{3,n}} \end{pmatrix}$$

and

$$A_4^{(n)} = \begin{pmatrix} \frac{c_{2,1}}{c_{3,1}} - k_d & & \\ & \ddots & \\ & & \frac{c_{2,n}}{c_{3,n}} - k_d \end{pmatrix}$$

with  $c_{i,j} := c_i(z_j)$  for all  $i = 1, \dots, 3$ , for all  $j = 1, \dots, n$ .

### 3.2.2 Considering a cone

As we want to regulate the state so that it converges to the equilibrium, one can easily see that this problem falls in the class described in Subsection 1.4.1. Indeed, we want to stabilize the linearized system so that the state variables are such that  $\tilde{S} \geq -\bar{S}$  and  $\tilde{X} \geq -\bar{X}$ . Considering the discretized system, we can introduce

$$\bar{x} := - \left( \bar{S}(z_1) \quad \cdots \quad \bar{S}(z_n) \quad \bar{X}(z_1) \quad \cdots \quad \bar{X}(z_n) \right)^T \in \mathbb{R}^{2n}$$

and we define the cone

$$C_{\bar{x}} := \{x \in \mathbb{R}^{2n} \mid x_i \geq \bar{x}_i, i = 1, \dots, 2n\}.$$

The problem is now to design a feedback matrix  $K$  that leads to the stability of the closed-loop system and to the invariance of the cone  $C_{\bar{x}}$ . However, regarding Theorem 1.4.2, the closed-loop matrix has to be Metzler in order for the cone  $C_{\bar{x}}$  to be invariant. This is impossible as  $A_2^{(n)}$  and  $A_3^{(n)}$  are nonpositive matrices, and  $b^{(n)}$  only acts on the first row of  $A^{(n)}$ . A possible way to address this issue is to add a new input to the system: the dilution rate. This is a standard way to achieve a better control of the system (see e.g. (Alwan 2012), (Henson 2010)). System (3.2.1)-(3.2.2) then becomes

$$\begin{cases} \frac{\partial S}{\partial t} &= D \frac{\partial^2 S}{\partial z^2} - (v + \theta) \frac{\partial S}{\partial z} - k\mu(S, X)X \\ \frac{\partial X}{\partial t} &= -k_d X + \mu(S, X)X \end{cases}$$

with boundary conditions

$$\begin{cases} D \frac{\partial S}{\partial z}(0, t) - (v + \theta)S(0, t) + (v + \theta)S_{in} &= 0 \\ \frac{\partial S}{\partial z}(L, t) &= 0 \end{cases}$$

for all  $t \geq 0$ , where  $\theta$  is the dilution rate. Linearizing and discretizing the system leads to a similar model as (3.2.5) except that

$$B^{(n)} = \begin{pmatrix} v & -\bar{S}(0) + \bar{S}_{in} - \frac{\partial \bar{S}}{\partial z}(z_1) \\ 0 & -\frac{\partial \bar{S}}{\partial z}(z_2) \\ \vdots & \vdots \\ 0 & -\frac{\partial \bar{S}}{\partial z}(z_n) \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{2n \times 2}$$

and

$$K^{(n)} = \begin{pmatrix} k_{1,1} & \cdots & k_{1,2n} \\ k_{2,1} & \cdots & k_{2,2n} \end{pmatrix} \in \mathbb{R}^{2 \times 2n}.$$

Obviously, it is still impossible to act on  $A_3$  and make it nonnegative, which seems logical as one cannot control the biomass directly but the substrate only. However, as  $A_3^{(n)}$  entries are relatively close to zero (around  $-0.0008$ ) one could approximate  $A_3^{(n)} = 0$  for the design of  $K^{(n)}$  only, then use Proposition 1.1.3 and solve the problem either by solving the matrix equation directly (see e.g. (Golub and Van Loan 1996, Algorithm 4.3.3)) or by parameterizing the solutions as in Section 2.2. The second method requires a bit of work and calculation but is actually feasible. First, one has to set conditions over  $K^{(n)}$  to ensure the Metzler property of the closed-loop system. Then, using Proposition 1.1.3, mathematical induction and the previously computed Metzler conditions in order to simplify the calculations, one can parameterize and rewrite a particular set of solutions for  $v$ . This leads to the  $\alpha$ -parameterization (with

$\alpha_n > 0$  arbitrary)

$$\begin{aligned} v_n &= \alpha_n \\ v_{n-1} &= \frac{-1}{p} \alpha_{n-1} - \frac{q_n}{p} \alpha_n \\ v_{n-i} &= \sum_{k=1}^{i+1} f_k \alpha_{n-i+k-1} \quad i = 2, \dots, n-1 \end{aligned}$$

with

$$\left\{ \begin{aligned} f_1 &= \frac{-1}{p} \\ f_2 &= \frac{q_{n-i+1}}{p^2} \\ f_j &= \frac{-r}{p} f_{j-2} + \frac{-q_{n-i+j-1}}{p} f_{j-1} \quad j = 3, \dots, i \\ f_{i+1} &= r f_{i-1} + q_n f_i \end{aligned} \right.$$

and where

$$\begin{aligned} p &= \frac{D_a}{\Delta z^2} + \frac{v}{\Delta z} \\ q_i &= \frac{-2D_a}{\Delta z^2} - \frac{v}{\Delta z} - k \frac{c_{1,i}}{c_{3,i}} \quad i = 2, \dots, n-1 \\ q_n &= \frac{-D_a}{\Delta z^2} - \frac{v}{\Delta z} - k \frac{c_{1,n}}{c_{3,n}} \\ r &= \frac{D_a}{\Delta z^2}. \end{aligned}$$

Although there exist solutions to this problem, one can show (analytically or numerically, the MATLAB routine that checks the feasibility of the problem can be found in Table 3.3) that the condition

$$(A^{(n)} + B^{(n)} K^{(n)}) \bar{x} \geq 0$$

is not compatible, implying that it is possible to positively stabilize the system considering a nonnegative initial condition, i.e. an initial condition that is higher than the equilibrium regarding the nominal system, but not to  $C_{\bar{x}}$ -invariably stabilize the system.



```

1 a = zeros(2*n^2-n,2*n);
2 b = zeros(2*n^2-n,1);
3 f = zeros(2*n,1);
4 for i = 1:n
5     a(i,:) = -x_bar';
6     b(i) = (A(i,:)*x_bar)/B(i,2);
7 end
8 for i = 1:n
9     test_a = -B(i,2)*eye(2*n,2*n);
10    test_a(i,:) = [];
11    a(n+(i-1)*(2*n-1)+1:n+i*(2*n-1),:) = test_a;
12    test_b = A(i,:);
13    test_b(i) = [];
14    b(n+(i-1)*(2*n-1)+1:n+i*(2*n-1)) = test_b;
15 end
16 [SOL,FVAL,EXITFLAG] = linprog(f,a,b);
17 BK = B(:,2)*SOL';
18 test = (A+BK)*x_bar;
19 for i = 1:2*n
20     if test(i) < 0
21         disp('Fail')
22     end
23 end

```

Table 3.3: Checking the feasibility of the cone invariance problem

**Result:** Figures 3.12 and 3.13 show the open-loop system trajectories, and one can easily see that it is unstable. Figures 3.14 and 3.15 show the closed-loop system trajectories that are obviously converging towards the equilibrium. They stay above the equilibrium at all time as we considered an initial condition  $x_0 \gg 0$  regarding the linearized system, i.e.  $x_0 \gg -\bar{x}$  regarding the nominal system. More precisely, we set  $n = 11$  and we choose

$$x_0 = \begin{pmatrix} \Delta z & \dots & \Delta z & 0.01\Delta z & \dots & 0.01\Delta z \end{pmatrix} \in \mathbb{R}^{22}$$

which is close enough to the equilibrium, a requirement when working with a linearized system to ensure the relevance of the obtained results. The MATLAB routines for the feedback design can be found in Tables 3.4 and 3.5.

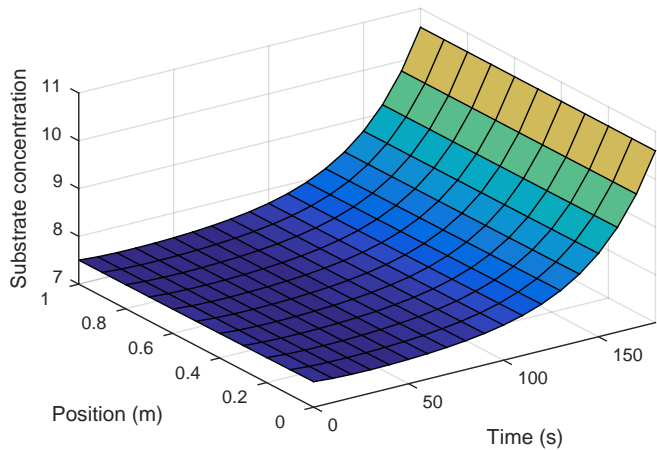


Figure 3.12: Open-loop substrate concentration trajectory ( $n = 11$ )

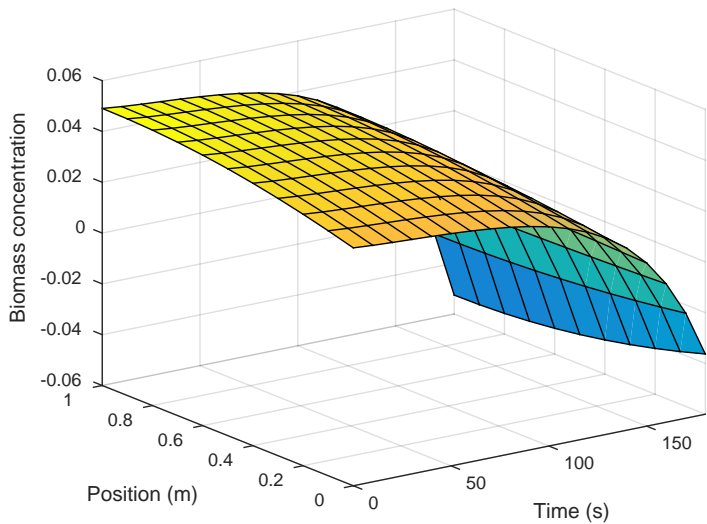


Figure 3.13: Open-loop biomass concentration trajectory ( $n = 11$ )

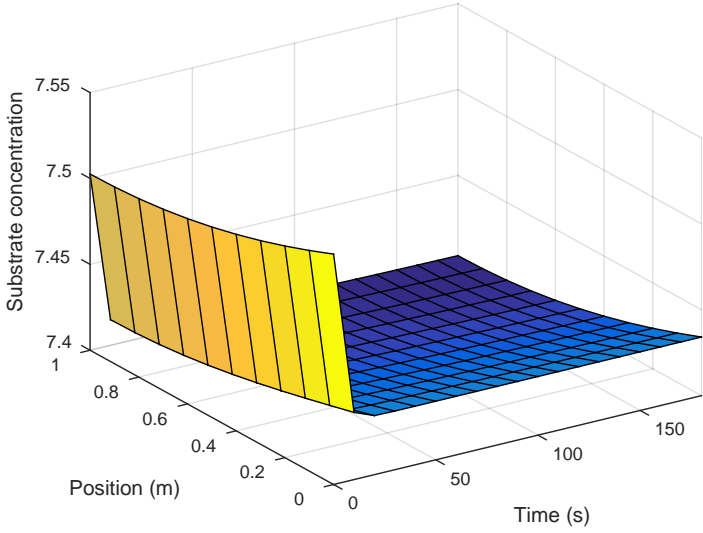


Figure 3.14: Closed-loop substrate concentration trajectory ( $n = 11$ )

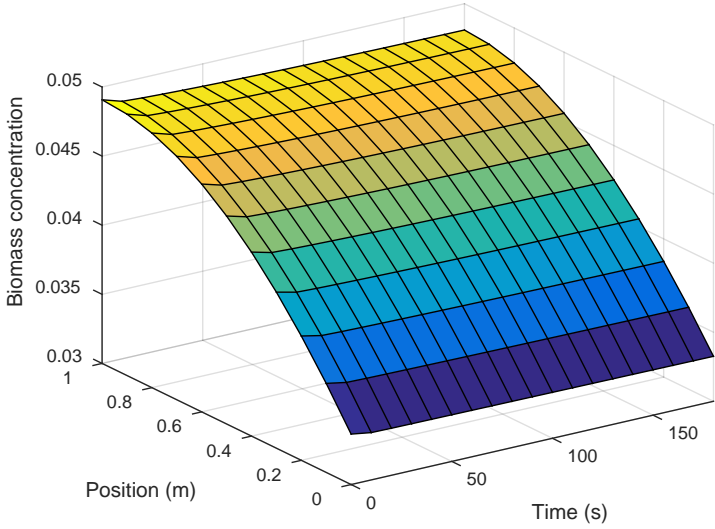


Figure 3.15: Closed-loop biomass concentration trajectory ( $n = 11$ )

```

1 % computing the vector v (see Golub p.153)
2 q = 2; % upper bandwidth
3 U = zeros(n,n); % upper band matrix
4 for i = 1:n-1
5     U(i,i) = -(D/Delta^2)+(nu/Delta));
6 end
7 U(n,n) = 1;
8 for i = 1:n-1
9     U(i,i+1) = (2*D)/(Delta^2) + (nu/Delta) + (k*c1(i+1))/c3(i+1);
10 end
11 for i = 1:n-2
12     U(i,i+2) = -(D/Delta^2);
13 end
14 Alpha = zeros(n,1);
15 for i = 1:n
16     Alpha(i) = 1;
17 end
18 for j = n:-1:1
19     Alpha(j) = Alpha(j)/U(j,j);
20     for i = max(1,j-q):j-1
21         Alpha(i) = Alpha(i) - U(i,j)*Alpha(j);
22     end
23 end
24 v = zeros(2*n,1);
25 v(1:n) = Alpha;

```

Table 3.4: Computation of  $v$  via (Golub and Van Loan 1996, Algorithm 4.3.3)

```

1 % computing the feedback K
2 omega = 10; % free parameter for K_11
3 K = zeros(2,2*n);
4 for i = n+2:2*n-1
5     K(2,i) = (k*(c2(i-n)/c3(i-n))) / (-(equi(i-n
        +1,1)-equi(i-n,1))/Delta);
6 end
7 K(2,2*n) = (k*(c2(n)/c3(n))) / (-(equi(n,1)-equi(n
    -1,1))/Delta);
8 K(1,1) = ((D/Delta^2)+(nu/Delta)+(k*(c1(1)/c3(1))
    )-(-(equi(1,1) + equi_input - ((equi(2,1)-equi
        (1,1))/Delta))*K(2,1) ) / nu) - omega;
9 K(1,2) = ((equi(1,1) - equi_input + ((equi(2,1)-
    equi(1,1))/Delta))/nu)*K(2,2) - (D/(nu*Delta^2)
    );
10 for i = 3:n-1
11     K(1,i) = ((equi(1,1) - equi_input + ((equi(2,1)-
        equi(1,1))/Delta))/nu)*K(2,i);
12 end
13 K(1,n) = ((equi(1,1) - equi_input + ((equi(n,1)-
    equi(n-1,1))/Delta))/nu)*K(2,n);
14 K(1,n+1) = ((equi(1,1) - equi_input + ((equi(2,1)-
    equi(1,1))/Delta))/nu)*K(2,n+1) + ((k*c2(1))/(
    nu*c3(1)));
15 for i = n+2:2*n-1
16     K(1,i) = ((equi(1,1) - equi_input + ((equi(2,1)-
        equi(1,1))/Delta))/nu)*K(2,i);
17 end
18 K(1,2*n) = ((equi(1,1) - equi_input + ((equi(2,1)-
    equi(1,1))/Delta))/nu)*K(2,2*n);

```

Table 3.5: Feedback design using  $v$ 

### 3.2.3 Considering a sector

As the Metzler property is obviously too strong, we can focus on the invariance of a sector around the equilibrium. This makes sense as we are working on a linearized system: we do not need all initial conditions to yield the set invariance, but only those that are "close" to the equilibrium. This allows to weaken the Metzler property, though one has to ensure that the state trajectory stays in the sector at all time which is a supplementary restriction: see Subsection 1.4.2, Theorem 1.4.4, and Corollaries 1.4.5 and 1.4.6. Unfortunately, it can actually be shown that these new conditions are also

too strong and thus unfeasible (the MATLAB routines that check the feasibility of the problem can be found in Tables 3.6 and 3.7), which means that one cannot expect to ensure the invariance of a given sector while stabilizing the system. One should also note that considering all the possible cases in Corollary 1.4.5, there are  $2^n$  conditions to verify which can be very costly as  $n$  grows when implemented as a routine.

```

1  % Using LMIs
2  tab={};
3  for i = 1:2*n
4      tab = cat(2,tab,x_double(i,:));
5  end
6  x_hat = allcomb(tab{:}); % preconceived function
7  Id = eye(2*n,2*n);
8  setlmis([]);
9  Q = lmivar(1,[2*n 1]);
10 Y = lmivar(2,[2 2*n]);
11 lmiterm([-1 1 1 Q],1,1); % Q
12 lmiterm([2 1 1 Q],A,1,'s'); % AQ+QA'
13 lmiterm([2 1 1 Y],B,1,'s'); % BY+Y'B'
14 for i = 1:length(y_hat)
15     for m = 1:2*n % Qy_hat >= x_hat + eps
16         lmiterm([- (2*i+1) 1 1 Q],diag(Id(m,:)),y_hat(i, :)'*Id(m,:));
17     end
18     lmiterm([- (2*i+1) 1 1 0],-diag(x_hat(i,:))');
19     lmiterm([- (2*i+1) 1 1 0],-eps*eye(2*n,2*n));
20     for m = 1:2*n % Qy_hat <= x_hat - eps
21         lmiterm([2*i+2 1 1 Q],diag(Id(m,:)),y_hat(i, :)'*Id(m,:));
22     end
23     lmiterm([2*i+2 1 1 0],-diag(x_hat(i,:))');
24     lmiterm([2*i+2 1 1 0],eps*eye(2*n,2*n));
25     disp(i);
26 end
27 Eigen_LMIs = getlmis;
28 [tmin,xfeas] = feasp(Eigen_LMIs);
29 Q = dec2mat(Eigen_LMIs,xfeas,Q);
30 Y = dec2mat(Eigen_LMIs,xfeas,Y);
31 K = Y/Q;

```

Table 3.6: Checking the feasibility of the sector invariance problem using LMIs

```

1 % Using the linprog function
2 tab = {};
3 for i = 1:2*n
4     tab = cat(2,tab,x_double(i,:));
5 end
6 x_hat = allcomb(tab{:}); % allcomb is a
    preconceived function
7 a = zeros(n*length(x_hat),2*n);
8 b = zeros(n*length(x_hat),1);
9 f = zeros(2*n,1);
10 for i = 1:n
11     for j = 1:length(x_hat)
12         if (x_hat(j,i) > 0)
13             a((i-1)*length(x_hat)+j,:) = x_hat(j,:);
14             b((i-1)*length(x_hat)+j) = (-A(i,:)*x_hat(
                j,:)')/B(i,2);
15         else
16             a((i-1)*length(x_hat)+j,:) = -x_hat(j,:);
17             b((i-1)*length(x_hat)+j) = -(-A(i,:)*x_hat(
                j,:)')/B(i,2);
18         end
19     end
20 end
21 [SOL,FVAL,EXITFLAG] = linprog(f,a,b);
22 BK = B(:,2)*SOL';

```

Table 3.7: Checking the feasibility of the cone invariance problem using *linprog*

**Result:** Although the sector invariance problem is not feasible for the biochemical reactor model, there is some interesting result that should be highlighted. As stated before, the inputs only act on the substrate as it would make no sense to control the biomass directly. This does not imply, however, that the sector invariance could not be achieved anyway for the biomass. Indeed, one can actually design the upper bound  $\bar{x}$  (not the lower bound  $\bar{x}$  as this one is fixed by nature) to ensure the sector invariance of the biomass. Considering the particular structure of the matrices  $A_3^{(n)}$  and  $A_4^{(n)}$ , one can see that the lower bound conditions are always verified. It is then sufficient to check the upper bound conditions. A fast calculation shows that the sector invariance

of the biomass is verified if  $\tilde{x}$  is such that

$$\tilde{x}_i < -\frac{\frac{c_{2,i}}{c_{3,i}} - k_d}{\frac{c_{1,i}}{c_{3,i}}} \bar{x}_{n+i} \quad \text{and} \quad \tilde{x}_{n+i} > -\frac{\frac{c_{1,i}}{c_{3,i}}}{\frac{c_{2,i}}{c_{3,i}} - k_d} \bar{x}_i$$

for  $i = 1, \dots, n$ . The fact that the arbitrary bounds can be used as parameters and tuned to achieve some desired properties should be taken into account while studying this kind of problems.

### 3.2.4 Considering Lyapunov level sets

In Subsection 3.2.2, it has been shown that Theorem 1.4.2 could not be applied on the biochemical reactor model in order to design the feedback matrix because the conditions were too strong. In Subsection 3.2.3 we restricted ourselves to a sector around the equilibrium, which implies we did not need all initial conditions in the nominal cone to yield the invariance. However, the state trajectory also had to stay in the sector at all time, which was an additional condition that also appeared to be too strong.

In this subsection, we go further as we want to find a suitable solution to the problem. In order to do so, let us consider a different approach. As it seems that asking the invariance of the cone or the sector for all initial conditions inside it is too strong, we first stabilize the system around the equilibrium (by use of a Lyapunov inequality for example, see Theorem 1.4.7), then compute a set of initial conditions around the equilibrium which guarantees that the state trajectory is in the cone or in the sector at all time. Again, this makes sense as we are working on a linearized model: limiting ourselves to a neighborhood of the equilibrium is totally relevant.

Regarding Subsection 1.4.3, we could define the Lyapunov function  $V(x) = x^T P x$ , using the positive definite symmetric matrix  $P$  from the Lyapunov inequality. By Theorem 1.4.10, we can compute

$$r = \min_{i \in \{1, \dots, n\}} \left\{ \frac{\det(P) \bar{x}_i^2}{\det(P^{[i]})} \right\}$$

which, we recall, is the smallest value such that  $V(x) = r$  is the equation of an ellipsoid that is tangent to (at least) one hyperplane, and not secant with any other hyperplane. This implies that the state trajectory remains in the ellipsoid (and thus in the cone) at all time, as long as we choose an initial condition that is such that  $x_0^T P x_0 \leq r$ . Let us set  $n = 11$  once again: using the pole placement function *place* in MATLAB, we compute a stabilizing feedback matrix

$$K := \begin{pmatrix} K_1 & K_2 \end{pmatrix} \in \mathbb{R}^{1 \times 22}$$

for the biochemical reactor system given by



$$K_1 = \begin{pmatrix} -1.9600 \\ -1.9598 \\ -1.9593 \\ -1.9585 \\ -1.9573 \\ -1.9557 \\ -1.9537 \\ -1.9512 \\ -1.9482 \\ -1.9447 \\ -1.9408 \end{pmatrix}^T \quad K_2 = \begin{pmatrix} 115.3195 \\ 114.8117 \\ 114.3331 \\ 113.8834 \\ 113.4621 \\ 113.0690 \\ 112.7038 \\ 112.3662 \\ 112.0560 \\ 111.7731 \\ 111.5172 \end{pmatrix}^T$$

and we compute (by use of LMIs) a positive definite symmetric matrix  $P$  such that  $(A+BK)^T P + P(A+BK)$  is a negative definite matrix. Note that one could rewrite the last condition as  $(A+BK)^T P + P(A+BK) = -\alpha I$  with  $\alpha > 0$  and use  $\alpha$  as a parameter to compute  $P$  in order to tune the shape of the level sets. Implementing Theorem 1.4.10 (the MATLAB routine can be found in Table 3.8), we obtain the radius  $r = 0.0061$ .

```

1  % computing K and P (Lyapunov matrix)
2  Poles = -abs(eig(A));
3  K = place(A,-B,Poles);
4  setlmis([]);
5  P = lmivar(1,[2*n 1]);
6  lmiterm([-1 1 1 P],1,1);
7  lmiterm([2 1 1 P],(A+B*K)',1,'s');
8  Eigen_LMIs = getlmis;
9  [tmin,xfeas] = feasp(Eigen_LMIs);
10 P = dec2mat(Eigen_LMIs,xfeas,P);
11
12 % computing r (contour line radius)
13 r_vec = zeros(2*n,1);
14 for i = 1:2*n
15     P_sub = P; P_sub(i,:) = []; P_sub(:,i) = [];
16     r_vec(i) = (det(P)*x_bar(i)^2)/det(P_sub);
17 end
18 r_Lyap = min(r_vec);

```

Table 3.8: Minimal Lyapunov level set radius computation

Choosing

$$x_0 = \begin{pmatrix} 0 & \dots & 0 & -0.05\Delta z & \dots & -0.05\Delta z \end{pmatrix} \in \mathbb{R}^{22}$$

leads to the state trajectories depicted in Figures 3.16 and 3.17. As the condition  $x_0^T P x_0 = 0.0025 \leq r$  is verified, one can expect the invariance of the corresponding level set to be guaranteed at all time, which implies that the state trajectory will remain in the cone.

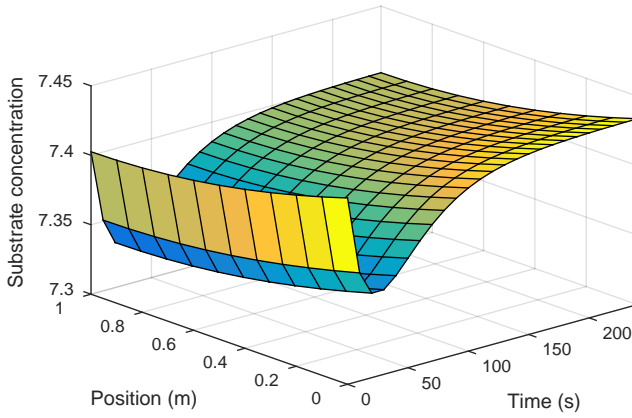


Figure 3.16: Closed-loop substrate concentration trajectory ( $n = 11$ )

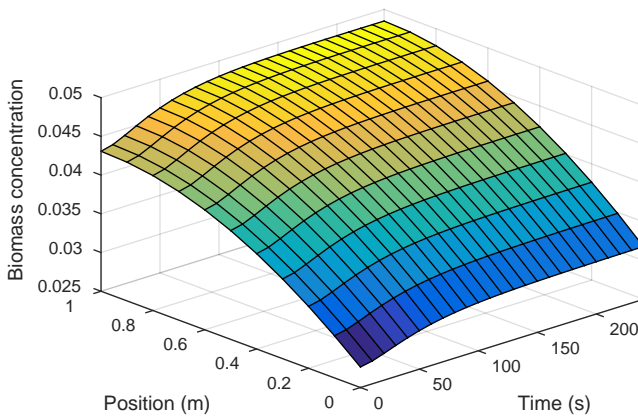


Figure 3.17: Closed-loop biomass concentration trajectory ( $n = 11$ )

Now choosing

$$x_0 = \begin{pmatrix} 20\Delta z & \cdots & 20\Delta z & -0.9\bar{x}_{n+1} & \cdots & -0.9\bar{x}_{2n} \end{pmatrix} \in \mathbb{R}^{22}$$

leads to the state trajectories depicted in Figures 3.18 and 3.19. This time we have  $x_0^T P x_0 = 1.2775 > r$ , which means we have no guarantee concerning the cone invariance. Fittingly, a quick check of Figure 3.19 shows that the biomass concentration does obviously not remain positive at all time.

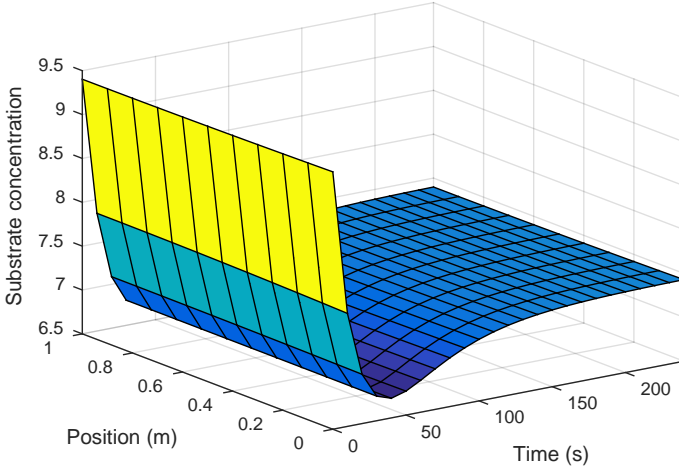


Figure 3.18: Closed-loop substrate concentration trajectory ( $n = 11$ )

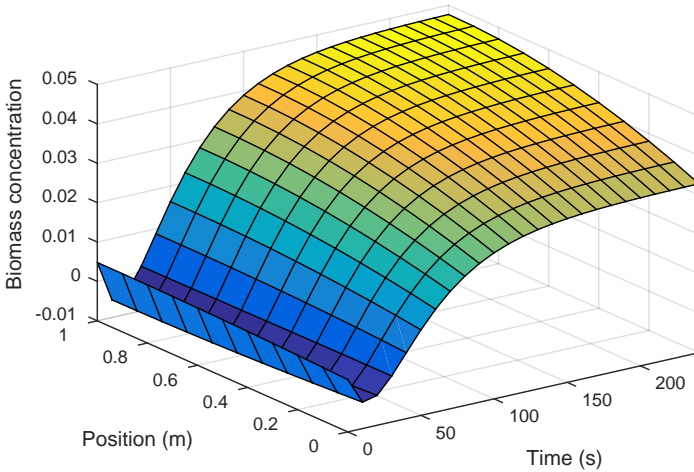


Figure 3.19: Closed-loop biomass concentration trajectory ( $n = 11$ )

Figure 3.20 focuses on a shorter time interval, thus bringing a better visibility of the invariance failure. In both cases, stabilization is achieved as expected, though, as the feedback matrix  $K$  was designed for that specific goal only.

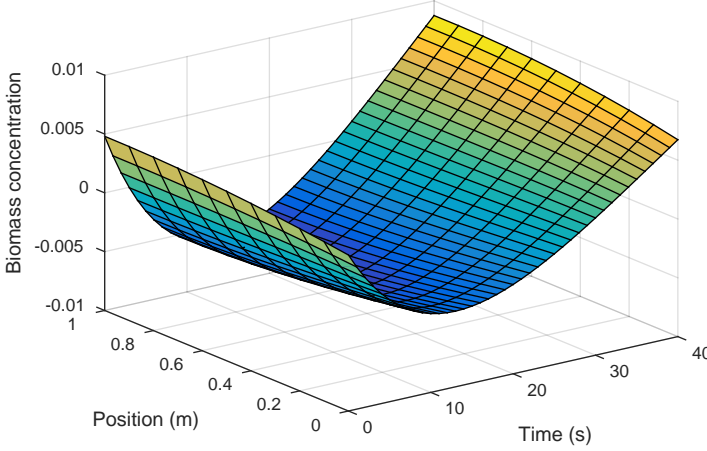


Figure 3.20: Closed-loop biomass concentration trajectory ( $n = 11$ )

So, we had to compromise to bring a solution to the biochemical reactor invariant stabilization problem as we could only ensure the invariance for a particular set of initial conditions. However, this method brings an interesting and elegant geometric approach, without any intricate calculation element, and that can be implemented as a fast and efficient numerical algorithm. Furthermore, as stated before, as we work on a linearized system it is relevant to focus on the equilibrium neighborhood only. It actually goes even further. Let us discretize the nominal nonlinear system (3.2.1)-(3.2.2). This leads to the nonlinear finite-dimensional model described by the equation

$$\dot{x}^{(n)} = \mathcal{A}^{(n)}x^{(n)} + \mathcal{E}^{(n)}f(x^{(n)}) + b^{(n)}u \quad (3.2.6)$$

where

$$\mathcal{A}^{(n)} = \begin{pmatrix} \mathcal{A}_1^{(n)} & 0 \\ 0 & \mathcal{A}_2^{(n)} \end{pmatrix} \in \mathbb{R}^{2n \times 2n},$$

$$\mathcal{E}^{(n)} = \begin{pmatrix} \mathcal{E}_1^{(n)} \\ \mathcal{E}_2^{(n)} \end{pmatrix} \in \mathbb{R}^{2n \times n},$$

$$b^{(n)} = \begin{pmatrix} v & 0 & \dots & 0 \end{pmatrix}^T \in \mathbb{R}^{2n},$$

$$x^{(n)} = \begin{pmatrix} S(z_1) & \dots & S(z_n) & X(z_1) & \dots & X(z_n) \end{pmatrix}^T \in \mathbb{R}^{2n}$$

and

$$\begin{aligned} f : \mathbb{R}^{2n} &\rightarrow \mathbb{R}^n \\ x^{(n)} &\mapsto \left( \mu(S(z_1), X(z_1))X(z_1) \quad \cdots \quad \mu(S(z_n), X(z_n))X(z_n) \right)^T \end{aligned}$$

and  $u$  is the input. The submatrices of  $\mathcal{A}^{(n)}$  and  $\mathcal{E}^{(n)}$  are given by

$$\mathcal{A}_1^{(n)} = \begin{pmatrix} \frac{-D}{\Delta z^2} - \frac{\nu}{\Delta z} & \frac{D}{\Delta z^2} & 0 & \cdots & 0 \\ \frac{D}{\Delta z^2} + \frac{\nu}{\Delta z} & \frac{-2D}{\Delta z^2} - \frac{\nu}{\Delta z} & \frac{D}{\Delta z^2} & & \vdots \\ 0 & & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \frac{D}{\Delta z^2} + \frac{\nu}{\Delta z} & \frac{-2D}{\Delta z^2} - \frac{\nu}{\Delta z} & \frac{D}{\Delta z^2} \\ 0 & \cdots & 0 & \frac{D}{\Delta z^2} + \frac{\nu}{\Delta z} & \frac{-D}{\Delta z^2} - \frac{\nu}{\Delta z} \end{pmatrix},$$

$$\mathcal{A}_2^{(n)} = -k_d I^{(n)}, \quad \mathcal{E}_1^{(n)} = -k I^{(n)} \quad \text{and} \quad \mathcal{E}_2^{(n)} = I^{(n)}.$$

One can easily see that linearizing this system would bring us back to system (3.2.5). This is not surprising as linearizing and discretizing the nominal nonlinear system should lead to the same model, regardless of the order in which we perform these two operations. Now let us define the following concept (see (Beauthier 2011, Definition 8.2.1), (Beauthier, Winkin and Dochain 2015, Definition 1)).

**Definition 15** *A nonlinear system is said to be locally positively state-invariant around the equilibrium  $x_e$  if there exists a neighborhood  $V_e$  of the equilibrium  $x_e$  such that for all  $x_0 \in V_e$  such that  $x_0 \gg 0$ , for all  $t \geq 0$ ,  $x(t) \gg 0$ .*

Then we state the following result (see (Beauthier 2011, Theorem 8.2.1), (Beauthier et al. 2015, Theorem 1)).

**Theorem 3.2.1** *If there exists a linear feedback control law  $K$  such that the linearized closed-loop system is stable and state-invariant with respect to the equilibrium  $x_e$ , i.e. such that*

$$\forall x_0 \in C_{\bar{x}}, \forall t \geq 0, \quad x(t) \in C_{\bar{x}}$$

*where  $\bar{x} = -x_e$  and  $x(t)$  is the solution of the linearized closed-loop system, then the resulting nonlinear closed-loop system is stable (the state trajectory converges to the equilibrium) and locally positively state-invariant around  $x_e$ .*

This actually implies that, as we managed to design a stabilizing feedback for the discretized linear system (3.2.5) that ensures the invariance of an ellipsoidal set inside the cone for some initial conditions, the properties also hold for the discretized

nonlinear system (3.2.6). One can then easily compute a feedback matrix that guarantees that the closed-loop state trajectory is stable and remains in a given cone for the discretized linear system, and be sure that these properties will be carried on when working on the discretized nonlinear system. However, one can hardly state anything about the nonlinear PDE system. It would then be interesting to check the convergence of the stabilizing feedback matrix to a stabilizing feedback operator for the nominal nonlinear system, as we did for the pure diffusion system in Section 2.3.



# Conclusion

*They say "doubt everything", but I disagree.  
Doubt is useful in small amounts,  
but too much of it leads to apathy and confusion.  
No, don't doubt everything. QUESTION everything.  
That's the real trick.*

– N. Sarabhai

In this work, a thorough analysis of the invariant stabilization problem of invariant systems was provided, dealing with the finite- and infinite-dimensional, linear and nonlinear cases.

As a first step we have shown with two formal proofs, one involving the state trajectory analysis and the other involving an extended system, why it is not possible to positively stabilize a positive system by use of a nonnegative input. In addition to providing an essential piece of information on the positive stabilization problem, it has served as a starting point for the design of positively stabilizing feedbacks, as it already gave an indication of the feedback matrix structure. Then naturally we have carried on and we have provided detailed methods enabling the parameterization of all positively stabilizing feedbacks for a particular class of positive finite-dimensional systems, namely single-input systems for which the input only acts on a single point of the domain. It has then been explained that it is actually possible to generalize this method to multiple-input systems for which the inputs act on any part of the domain, though the problem would be a lot more complicated to solve and would probably require some alterations and adjustments in order to obtain a refined and practical result as for the single-input systems case. Once these feedbacks have been designed, it has been envisaged to go further in the positive stabilization problem by considering a more general form of invariance. We then have introduced the concepts of cone invariance (positivity being a particular case obviously, as it represents the invariance of the nonnegative orthant) and we have extended this concept to that of sector invariance, by adding upper bounds to the state. From then on, it has been possible to provide some necessary and sufficient conditions that guarantee the invariance of such a subset. Finally, we have developed a complementary result for the invariance of cones or sectors, by exploiting the level sets of Lyapunov functions. Although this method



restrains us to particular sets of initial conditions, it is far more flexible than the previous ones and subsequently guarantees finding a non-empty set of initial conditions for which we achieve both stabilization and invariance. As the set can potentially be restricted to a small neighborhood around the equilibrium, this method is particularly adapted to linearized systems, for which it is relevant to work only locally around the equilibrium.

As the previous results only apply to the case of linear finite-dimensional systems, we have then looked into the case of linear infinite-dimensional systems. First, we have shown why and how considering a nonnegative input could actually lead to the positive stabilization of a positive system. When dealing with a PDE system where the input acts in the boundary conditions, it is shown that it is possible to design an equivalent expression of the system where the input acts in the dynamics. According to the class of systems, the input change can consist in a simple sign inversion (with intervention of a Dirac delta distribution), which means that a nonnegative input acting in the boundary conditions is equivalent to a nonpositive input acting in the dynamics and would thus make the positive stabilization of the system possible. We have then studied the pure diffusion system with Neumann boundary conditions, considering both cases where the input acts in the boundary conditions or in the dynamics, as we have discretized the system and applied the previous theoretical results. So, we have provided a finite-dimensional model, using the finite difference method, that has the same main features as the nominal system. We then have been able to parameterize all the positively stabilizing feedbacks for the pure diffusion model, along with three different and peculiar subparameterizations to complete the main result, we have performed some numerical simulations to support the theory, and we have shown the convergence of the positively stabilizing feedback matrix to a positively stabilizing feedback operator for the PDE system, as well as the convergence of the closed-loop system to the closed-loop PDE system. To achieve this, we have carried out a study of the numerical scheme, highlighting its consistency and its stability and thus its convergence, as well as considered a state space approach. Finally, we have discussed a question that has arisen during the realization of the thesis, namely the status of the boundary conditions when we discretize a PDE system, as they do not intervene in the finite-dimensional models. Thus we have provided a methodology that allows the positive stabilization of positive systems while ensuring that the boundary conditions, that have also been discretized (as algebraic conditions) and converge to the nominal boundary conditions, are verified at all time.

Finally, we have gone one more step further and have dealt with the case of positive nonlinear systems. We have explained how this class of problems can be reduced to cone-invariant stabilization problems of positive linear systems that subsequently fall in the class of systems that were studied previously. We have then analyzed a technical and relevant application, namely a tubular biochemical reactor model, for which we have described the model notably by performing an equilibria study that showed us towards which "optimal" profile we wish the state to converge in order to maximize the biomass production. Once this has been done, we have linearized

the system around the desired equilibrium, then we have discretized the system in a manner similar to that used for the pure diffusion system, using the finite difference method. Thus we have been able to deal with the invariant stabilization problem of a cone, which appeared not to be solvable for this particular model as the conditions required for the cone invariance, or even for a sector invariance, have proved to be too strong. We have then applied the results concerning Lyapunov functions level sets, that made it possible to compute a set of initial conditions guaranteeing the invariance of ellipsoidal subsets post-stabilization, all of which has been numerically illustrated via simulations. Eventually, we have concluded this chapter by explaining that a positively stabilizing feedback for the linearized system is certainly positively stabilizing for the nonlinear system (locally), which makes the previous results even more interesting as they allow a quick and easy design of an invariably stabilizing feedback for the nominal nonlinear system.

Numerous routines have been coded in MATLAB in order to perform the numerical simulations for the pure diffusion system and the tubular biochemical reactor model. The codes of the main routines can be found in the appendix.

## Perspectives

When dealing with a general problem such as the invariant stabilization, for multiple classes of systems, analytically and numerically, this opens a lot of doors and it is subsequently always natural to consider further investigations that could reveal to be very interesting and might complete the results that have been presented throughout this work.

As mentioned previously, when designing the positively stabilizing feedbacks we have considered a particular class of positive systems. It would be then interesting to generalize these results to any type of positive system. It would also be interesting to add some optimization aspect in the feedback design, i.e. to optimize the choice of a positively stabilizing feedback in the admissible set with respect to some given criterion.

Corollary 1.4.6 provides simple conditions for the invariance of a sector. However, as previously stated, it is hardly practical in the current state as it requires information that is only known after stabilizing the system. It could be useful to develop the result and go further in order to check if it is possible to make it practical.

Lyapunov functions level sets have been deeply analyzed so that we obtained a result that is geometrically appealing, simple enough and very efficient regarding the computation time. However, studying the choice of the positive definite matrix  $P$ , or even of other Lyapunov functions, could allow to tune the level sets so that they match the considered cone or sector better. This way, one could improve the design of the set of initial conditions.

The results described in (Emirsjlow and Townley 2000) concerning the change of input acting in the boundary conditions to an input acting in the dynamics are presented via particular examples only, namely the Heat equation and the Wave equation. It would be opportune to extend the results to entire classes of systems, e.g. to convection-diffusion-reaction systems.

One could also find conditions over any discretized feedback so that it converges to a positively stabilizing feedback for the nominal PDE system.

The question of the discretized boundary conditions preservation merits some more attention, notably regarding the relevance or necessity of this kind of additional conditions. As mentioned previously, it is possible to consider weaker conditions for the discretized system, that would guarantee anyway that the boundary conditions of the nominal system are verified when the discretization step tends to zero.

Chapter 3 shows via an application how to design a positively regulating feedback for a positive nonlinear system by using a linearized and discretized model. It would be interesting to develop and study the nonlinear closed-loop system, by applying the feedback on the discretized nonlinear model, by performing numerical simulations, and by studying the convergence of both the feedback and the resulting closed-loop system, as it has been done for the pure diffusion system.

Finally, we can only encourage people to continue to develop and analyze relevant and trendy applications, as this work notably aims to provide practical tools for this kind of systems. There is no doubt that new questions would arise and it would be very useful to extend and adapt the results of this work to interesting applications such as, for example, population dynamics models, pharmacokinetics, renewable energy devices and many others.

# Bibliography

- B. Abouzaid, J.J. Winkin, and V. Wertz. Positive stabilization of infinite-dimensional linear systems. *in* 'Proceedings of the 49th IEEE Conference on Decision and Control (CDC)', pp. 845–850, 2010.
- M.E. Achhab and J.J. Winkin. Stabilization of infinite dimensional systems by state feedback with positivity constraints. *in* 'Mathematical Theory of Networks and Systems (MTNS), 2014 21st International Symposium on', 2014.
- M.E. Achhab and J.J. Winkin. Positive stabilization of a class of infinite-dimensional positive systems. *in* F. Cacace, L. Farina, R. Setola and A. Germani, eds, 'Positive Systems: Theory and Applications (POSTA 2016)', pp. 191–200. Springer International Publishing, 2017.
- M. Ait Rami. Solvability of static output-feedback stabilization for LTI positive systems. *Systems & Control Letters*, **60**, 704–708, 2011.
- M. Ait Rami and D. Napp. Characterization and stability of autonomous positive descriptor systems. *IEEE Transactions on Automatic Control*, **57**(10), 2668–2673, 2012.
- I. Aksikas, J.J. Winkin, and D. Dochain. Optimal LQ-feedback regulation of a non-isothermal plug flow reactor model by spectral factorization. *IEEE Transactions on Automatic Control*, **52**(7), 1179–1193, 2007.
- G. Alwan. Simulation and optimization of a continuous biochemical reactor. *Chemical and Process Engineering Research*, **5**, 2012.
- K.J. Arrow. A "dynamic" proof of the Frobenius-Perron theorem for Metzler matrices. *Technical Report Series*, **542**, 1989.
- J. Banasiak and L. Arlotti. *Perturbations of Positive Semigroups with Applications*. Springer Monographs in Mathematics. Springer London, 2006.
- A. Bátkai, M.K. Fijavž, and A. Rhandi. *Positive Operator Semigroups: From Finite to Infinite Dimensions*. Operator Theory: Advances and Applications. Springer International Publishing, 2017.

- C. Beauthier. *The LQ-Optimal Control Problem for Invariant Linear Systems*. PhD thesis, University of Namur, 2011.
- C. Beauthier, J.J. Winkin, and D. Dochain. Input/state invariant LQ-optimal control: Application to competitive coexistence in a chemostat. *Evolution Equations and Control Theory*, **4**(2), 143–158, 2015.
- A. Berman, M. Neumann, and R.J. Stern. *Nonnegative matrices in dynamic systems*. Pure and applied mathematics. Wiley, 1989.
- F. Blanchini. Set invariance in control. *Automatica*, **35**(11), 1747–1767, 1999.
- F. Blanchini and S. Miani. *Set-Theoretic Methods in Control*. Systems & Control: Foundations & Applications. Birkhäuser Boston, 2007.
- S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. Society for Industrial and Applied Mathematics (SIAM), 1994.
- A.M. Bruckner, J.B. Bruckner, and B.S. Thomson. *Real Analysis*. Prentice-Hall, 1997.
- J. B. Burl. Singular system - A tutorial. in ‘Nineteenth Asilomar Conference on Circuits, Systems and Computers, 1985.’, pp. 529–533, 1985.
- E.B. Castelan and J.C. Hennet. On invariant polyhedra of continuous-time linear systems. *IEEE Transactions on Automatic Control*, **38**(11), 1680–1685, 1993.
- V. Chellaboina, S.P. Bhat, W.M. Haddad, and D.S. Bernstein. Modeling and analysis of mass-action kinetics: Nonnegativity, realizability, reducibility, and semistability. *IEEE Control Systems Magazine*, **29**(4), 60–78, 2009.
- P.D. Christofides and P. Daoutidis. Nonlinear feedback control of parabolic PDE systems. in R. Berber and C. Kravaris, eds, ‘Nonlinear Model Based Process Control’, pp. 371–399. Springer Netherlands, Dordrecht, 1998.
- R. Curtain and K. Morris. Transfer functions of distributed parameter systems: A tutorial. *Automatica*, **45**(5), 1101–1116, 2009.
- R.F. Curtain and H.J. Zwart. *An Introduction to Infinite-Dimensional Linear Systems Theory*. Springer-Verlag, 1995.
- L. Dai. *Singular Control Systems*. Lecture Notes in Control and Information Sciences. Springer Berlin Heidelberg, 1989.
- P. De Leenheer and D. Aeyels. Stabilization of positive linear systems. *Systems & Control Letters*, **44**, 259–271, 2001.
- J.N. Dehaye. Stabilisation positive de systèmes différentiels linéaires : théorie et application aux réacteurs tubulaires. Master’s thesis, University of Namur, 2011.

- J.N. Dehaye and J.J. Winkin. Parameterization of positively stabilizing feedbacks for single-input positive systems. *Systems & Control Letters*, **98**, 57–64, 2016.
- J.N. Dehaye and J.J. Winkin. Positive stabilization of a diffusion system by non-negative boundary control. in F. Cacace, L. Farina, R. Setola and A. Germani, eds, ‘Positive Systems: Theory and Applications (POSTA 2016)’, pp. 179–190. Springer International Publishing, 2017.
- C. Delattre, D. Dochain, and J. Winkin. Sturm-Liouville systems are Riesz-spectral systems. *Int. J. Appl. Math. Comput. Sci.*, **13**(4), 481–484, 2003.
- A.K. Dramé, D. Dochain, and J.J. Winkin. Asymptotic behavior and stability for solutions of a biochemical reactor distributed parameter model. *IEEE Transactions on Automatic Control*, **53**, 412–416, 2008.
- Z. Emirsjlow and S. Townley. From PDEs with boundary control to the abstract state equation with an unbounded input operator: A tutorial. *European Journal of Control*, **6**, 27–49, 2000.
- L. Farina and S. Rinaldi. *Positive Linear Systems: Theory and Applications*. John Wiley & Sons, 2000.
- P. Frey. The finite difference method. <https://www.ljll.math.upmc.fr/frey/ftp/finite-differences.pdf>, 2009. [Online; accessed 2-August-2017].
- G.H. Golub and C.F. Van Loan. *Matrix Computations (Third Edition)*. The Johns Hopkins University Press, 1996.
- W.M. Haddad, V. Chellaboina, and Q. Hui. *Nonnegative and Compartmental Dynamical Systems*. Princeton University Press, 2010.
- P. Hartman. On invariant sets and on a theorem of Wazewski. *Proceedings of the American Mathematical Society*, **32**(2), 511–520, 1972.
- M. A. Henson. Model-based control of biochemical reactors. in W. Levine, ed., ‘The Control Handbook, Second Edition: Control System Applications, Second Edition’, chapter 19. CRC Press, 2010.
- R.A. Horn and C.R. Johnson. *Matrix Analysis*, Vol. 1. Cambridge University Press, 1990.
- T. Kaczorek. *Positive 1D and 2D Systems*. Springer-Verlag London, 2002.
- I. Karafyllis and M. Krstic. Stability of integral delay equations and stabilization of age-structured models. *CoRR*, **abs/1609.09437**, 2016.
- H.K. Khalil. *Nonlinear Systems*. Pearson Education. Prentice Hall, 2002.
- M.G. Krein and M.A. Rutman. Linear operators leaving invariant a cone in a Banach space. *Uspekhi Mat. Nauk*, **3:1**(23), 3–95, 1948.

- M. Laabissi, M.E. Achhab, J.J. Winkin, and D. Dochain. Trajectory analysis of non-isothermal tubular reactor nonlinear models. *Systems & Control Letters*, **42**, 169–184, 2001.
- M. Laabissi, J.J. Winkin, D. Dochain, and M.E. Achhab. Dynamical analysis of a tubular biochemical reactor infinite-dimensional nonlinear model. in ‘Proceedings 44th IEEE CDC-ECC’, pp. 5965–5970, 2005.
- Jr. Lasley, J.W. On degenerate conics. *The American Mathematical Monthly*, **64**(5), 362–364, 1957.
- F.L. Lewis. A survey of linear singular systems. *Circuits, Systems and Signal Processing*, **5**(1), 3–36, 1986.
- F. Logist, P. Saucez, J. Van Impe, and A. Vande Wouwer. Simulation of (bio) chemical processes with distributed parameters using Matlab (R). *The Chemical Engineering Journal*, **155**(3), 603–616, 2009.
- B. Roszak and E.J. Davison. Necessary and sufficient conditions for stabilizability of positive LTI systems. *Systems & Control Letters*, **58**, 474–481, 2009.
- B.S. Rüffer. Monotone inequalities, dynamical systems, and paths in the positive orthant of Euclidean  $n$ -space. *Positivity*, **14**(2), 257–283, 2010.
- S.H. Saperstone. Global controllability of linear systems with positive controls. *SIAM Journal on Control*, **11**(3), 417–423, 1973.
- A. Schrijver. *Theory of Linear and Integer Programming*. John Wiley & Sons, 1998.
- B. Shafai and C. Li. Positive stabilization of linear continuous-time singular systems. in ‘2016 World Automation Congress (WAC)’, pp. 1–7, 2016.
- H.L. Smith. *Monotone Dynamical Systems: An Introduction to the Theory of Competitive and Cooperative Systems*. American Mathematical Soc., 2008.
- H.L. Smith and X.-Q. Zhao. Dynamics of a periodically pulsed bio-reactor model. *Journal of Differential Equations*, **155**(2), 368–404, 1999.
- J.W. Thomas. *Numerical Partial Differential Equations: Finite Difference Methods*. Springer, 1995.
- A. Vande Wouwer, P. Saucez, and C.V. Fernández. *Simulation of ODE/PDE Models with MATLAB®, OCTAVE and SCILAB: Scientific and Engineering Applications*. SpringerLink : Bücher. Springer International Publishing, 2014.
- J.M. Varah. A lower bound for the smallest singular value of a matrix. *Linear Algebra and its Applications*, **11**(1), 3–5, 1975.
- G. Verghese, B. Levy, and T. Kailath. A generalized state-space for singular systems. *IEEE Transactions on Automatic Control*, **26**(4), 811–831, 1981.

- E. Virnik. Stability analysis of positive descriptor systems. *Linear Algebra and its Applications*, **429**(10), 2640–2659, 2008.
- F. Vivien. On the optimality of Feautrier’s scheduling algorithm. in B. Monien and R. Feldmann, eds, ‘Euro-Par 2002 Parallel Processing: 8th International Euro-Par Conference Paderborn’, pp. 299–309. Springer-Verlag, 2002.
- H.-H. Wang, M. Krstić, and G. Bastin. Optimizing bioreactors by extremum seeking. *International Journal of Adaptive Control and Signal Processing*, **13**(8), 651–669, 1999.
- J.J. Winkin, D. Dochain, and P. Ligarius. Dynamical analysis of distributed parameter tubular reactors. *Automatica*, **36**, 349–361, 2000.
- L. Zhang and C. Prieur. Necessary and sufficient conditions on the exponential stability of positive hyperbolic systems. *IEEE Transactions on Automatic Control*, **62**(7), 3610–3617, 2017.



